Molecular dissection of PrimPol, a novel primase-polymerase involved in damage tolerance during DNA replication in eukaryotic cells

A thesis submitted to the University of Sussex for the degree of Doctor of Philosophy

Benjamin A. Keen

March 2015

I hereby declare that this thesis has not been, and will not be, submitted in whole or in part to another university for the award of any other degree.

Benjamin A. Keen

University of Sussex

Benjamin A. Keen

Doctor of Philosophy Biochemistry

Molecular dissection of PrimPol, a novel primase-polymerase involved in damage tolerance during DNA replication in eukaryotic cells

Summary

PrimPol is a recently identified member of the archaeo-eukaryotic primase (AEP) family of proteins. It possesses both primase and polymerase activities and is involved in the replication of both nuclear and mitochondrial DNA. PrimPol is predicted to possess an AEP polymerase and a UL52-like zinc finger domain. This thesis establishes the roles of these domains in the context of PrimPol's catalytic activities. Although apparently dispensable for polymerase activity, the zinc finger is essential for maintaining primase activity and also appears to play an important role in regulating the processivity and fidelity of PrimPol's extension activities. A recently study identified a PrimPol mutation (Y89D) that is potentially associated with the development of high myopia in humans. Here, the biochemical defects associated with this mutant are analysed and described. This protein variant has a significant reduction in polymerase activity. Mutational analysis suggests that the hydrophobic ring of tyrosine is important for retaining wild-type DNA extension activity. Biophysical analysis of the secondary structure and stability of this PrimPol variant suggests that this PrimPol variant has reduced α-helical content and is less stable than the wild-type protein.

Finally, the interaction of PrimPol with single-stranded DNA binding protein replication protein A (RPA) is investigated. Previous studies have identified an interaction of PrimPol with RPA. Here, it is demonstrated that PrimPol has two separate RPA interaction motifs and a crystal structure is presented of one such motif in PrimPol bound to RPA that reveals the molecular basis for this interaction.

Together, these studies provide molecular insights into the catalytic mechanism of PrimPol as well as some of the key intramolecular and intermolecular mechanisms of that regulate the activities of PrimPol.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Aidan Doherty, for the opportunity to carry out a doctorate in his laboratory on such an exciting project and for all the assistance and guidance he has offered me these past few years. I would also like to thank all of the members of the Doherty lab during my tenure, including Nigel, Stanislaw, Laura, Julie, Edward, Sean, Tom, Pierre, Helen, Alfredo, Farimah, Violetta and Przemek for helping me find my feet in the laboratory and their continued help and ideas.

Special thanks go to Stanislaw, who taught me a lot of the biochemical techniques I've used throughout this thesis, Nigel, who is responsible for a lot of my understanding about protein purification, crystallography and many of the biophysical techniques I have used, and Laura, who validated a lot of the results shown here *in vivo*.

I'd also like to acknowledge Chris Dadswell, for his assistance with inductively coupled plasma mass spectrometry, Sarah Parry-Morris and Owen Wells, for their help with circular dichroism studies, and Antony Oliver, for his help with thermal denaturation and fluorescence anisotropy. My gratitude is also due to Prof. Shigenori Iwai who kindly provided DNA containing a 6–4 photoproduct used in studies presented here.

My research was supported by the British Biotechnology and Biological Sciences Research Council and the University of Sussex and they, too, both have my thanks.

Finally I'd like to thank everyone in the Genome Damage and Stability Centre for their continued support, friendly conversations, and for providing an excellent environment to work.

Table of Contents

Abbreviations

(6-4)PP	Pyrimidine (6-4) pyrimidone photoproducts
3meA	N3-methyladenine
7meG	N7-methylguanine
8-oxo-G	8-oxo-2’-deoxyguanosine
AEP	Archaeo-eukaryotic primase
AMP-PNP	Adenylyl imidodiphosphate
AP	Apurinic/apyrimidinic
APE1	AP endonuclease 1
APLF	Aprataxin-and-PNK-like factor
APS	Ammonium persulphate
AT	Ataxia telangiectasia
ATP	Adeonsine triphosphate
ATR	AT and Rad3-related kinase
A_x	Absorbance at wavelength x nanometres
BER	Base excision repair
bp	Base pairs
BPDE	Benzo[a]pyrenediolepoxide
BSA	Bovine serum albumin
CAK	Cyclin-dependent kinase-activating kinase
CD	Circular dichroism
CDC	Cell division cycle
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
CDT1	CDC10 Target 1
CldU	5-chloro-2'-deoxyuridine
CMG	CDC45, MCM, GINS
CPD	Cyclobutane pyrimidine dimer
CS	Cockayne syndrome
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
DDK	Dfb4-dependent kinase
dGTP	Deoxyguanosine triphosphate

DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic Acid
dNMP	Deoxynucleoside monophosphate
dNTP	Deoxynucleoside triphosphate
DSB	Double-stranded DNA break
DSBR	Double strand break repair
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
dUTP	Deoxyuridine triphosphate
EMSA	Electrophoretic mobility shift assay
exo	Exonuclease
FAD	Flavin adenine dinucleotide
FEN1	Flap Endonuclease 1
HMG	High mobility group
HR	Homologous recombination
HRP	Horseradish peroxidase
HSQC	Heteronuclear single quantum coherence
ICP-MS	Inductively coupled plasma mass spectrometry
IdU	5-iodo-2'-deoxyuridine
IMAC	Immobilised metal affinity column
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LC-MS/MS	Liquid chromatography tandem mass spectrometry
MCM	Mini-chromosome maintenance
MGMT	O6-methylguanine-DNA-methyltransferase
MMEJ	Microhomology-mediated end joining
MMR	Mismatch Repair
MPD	2-methyl-2,4-pentanediol
MRN	Mre11, Rad50, Nbs1
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
mtSSB	Mitochondrial single-stranded binding protein

NCLDV	Nucleo-cytoplasmic large DNA virus
NCR	Non-coding region
NER	Nucleotide Excision Repair
NHEJ	Non-homologous end-joining
Ni-NTA	Nickel-nitrolotriacetic acid
NMR	Nuclear magnetic resonance
OB	Oligosaccharide/Oligonucleotide-binding
OD_x	Optical density at wavelength x nanometres
ORC	Origin of replication complex
ORF	Open reading frame
PAD	Polymerase associated domain
PARP1	poly(ADP-ribose) polymerase 1
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PCT	Pre-Crystallization Test
PDB	Protein data bank
PEO	Progressive External Ophthalmoplegia
pI	Isoelectric point
PMSF	Phenylmethanesulfonylfluoride
PNKP	Polynucleotide kinase/phosphatase
pol	Polymerase
polRMT	Mitochondrial DNA-directed RNA polymerase
pre-RC	Pre-replication complex
PSI-BLAST	Position-specific iterative basic local alignment search tool
PVDF	Polyvinyladine fluoride
RBM	RPA-binding motif
RFC	Replication factor C
RMSD	Root-mean-square-deviation
RNA	Ribonucleic Acid
rNTP	Ribonucleoside triphosphate
RPA	Replication protein A
rpm	Revolutions per minute
SDM	Site-directed mutagenesis

SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SDSA	Synthesis-dependent strand annealing
SEC	Size-exclusion chromatography
SOB	Super optimal broth
SSB	Single-stranded DNA break
ssDNA	Single-stranded DNA
TB	Terrific broth
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFAM	Mitochondrial transcription factor A
TFIIH	Transcription factor II Human
Tg	Thymine glycol
TLS	Translesion DNA synthesis
Topo	Topoisomerase
tRNA	Transfer RNA
UV	Ultraviolet
XLF	XRCC4-like factor
XP	Xeroderma pigmentosum
XRCC4	X-ray repair cross-complementing protein 4

N.B. In addition to the above abbreviations, standard SI units, periodic table abbreviations, amino acid three-letter and single-letter abbreviations and DNA base single-letter abbreviations are used.

List of Figures

List of Tables

CHAPTER 1

Introduction

1. Introduction

This thesis is introduced by the experimentally tested scientific theories that preceded the undertaking of the work presented here. The topics covered include the replication of the genome, damage to the genome and how it is managed, the replication of DNA through damaged genome components, polymerases that synthesise novel DNA strands including the archeao-eukaryotic primase (AEP) superfamily of proteins, and a protein of this superfamily, PrimPol (previously named coiled-coil domain containing protein 111, CCDC111, FLJ33167). PrimPol is an AEP that is found in humans and other eukaryotes and is the main subject of this investigation.

1.1. The Genetic Code

DNA is the genetic blueprint for life on Earth; within this impressive molecule lies the information required for the development, functioning and survival of a diverse array of living organisms (Avery et al., 1944; Griffith, 1928; Hershey and Chase, 1952). In 1953, using the X-Ray data of Wilkins and Franklin and information of the base ratio data from Chargaff, Watson and Crick published their ground-breaking discovery of the structure of DNA (Chargaff et al., 1952; Watson and Crick, 1953; Zamenhof et al., 1952). DNA is a double helical molecule, with two anti-parallel strands of polynucleotides, each consisting of nucleotide monomers composed of a deoxyribose-phosphate backbone and an inward-facing perpendicular nitrogenous base. Phosphodiester bonds join the monomers and the two strands are held together by non-covalent hydrogen bonds that are complementary between adenine and thymine, and guanine and cytosine. Watson and Crick had hypothesised, based on their double-helical model of DNA, that replication occurs in a semi-conservative fashion i.e. novel copies of DNA contain one daughter strand and one parent strand, which was subsequently proved correct (Meselson and Stahl, 1958). Experiments in the following years determined how DNA provided the genetic code for the production of proteins. The DNA is transcribed into mRNA, which is subsequently translated to produce proteins (Crick, 1970). Three bases, known as a ‘codon’, code one amino acid and the following three bases code the next amino acids; these codons do not overlap (Crick et al., 1961; Tsugita and Fraenkel-Conrat, 1960). The Nobel Prize for “cracking the genetic code” was ultimately awarded to Khorana and Nirenberg, who were responsible for mapping each of the codons to the amino acids that they code (Nirenberg et al., 1965).

1.2. Genome Replication

Replication of a genome is a herculean task. Each time a human cell divides it needs to copy ~3 x 10^9 base pairs at a rate of ~1,200 bp min^{-1} per replication fork, and it must do so faithfully else risk permanent changes in the genetic information that can pass on to daughter cells and, potentially, lead to deregulated metabolic processes and cancer (Lebofsky et al., 2006). Maintaining genomic integrity is a fundamental aspect of life and requires a specialised, highly tuned machinery to do so. The replication of the genome occurs in three stages: initiation, elongation and termination.

1.2.1. Origins of Replication – Where to Begin

Given the rate of replication in human cells, if there were only one origin of replication on each chromosome it would take around 70 days to replicate the entire genome. As such, there are multiple origins of replication in each cell and, in mammals, this number of these parallel synthesis sites is around 30,000-50,000 (Cairns, 1966; Huberman and Riggs, 1968; 1966). There are three steps in the setting of replication origins, which include recognition of origins, the assembly of a pre-replication complex (pre-RC) and the activation of the pre-RC (Figure 1.1). The pre-RC is assembled in G1 phase (Rao and R. T. Johnson, 1970). The origin of replication complex (ORC) is a heterohexamer with DNA-dependent ATPase activity (S. P. Bell and Stillman, 1992). Despite having a consensus sequence in *Saccharomyces cerevisiae* (reviewed in (Newlon and Theis, 1993)), there is no consensus sequence for metazoan origins of replication (Vashee et al., 2003). Following ORC binding to an origin, two factors, known as cell divison cycle 6 (CDC6) and CDC10 Target 1 (CDT1) are recruited to the origin and load the mini-chromosome maintenance (MCM) helicase complex MCM2-7 in an ATP-dependent manner in a process known as licensing (Remus et al., 2009). MCM9 also works in concert with CDT1 to load the MCM2-7 complex (Lutzmann and Méchali, 2008). Licensing marks the end of the pre-RC complex assembly and the complex is ready to be activated. Two kinases that are S phase-specific are required prior to the activation of the pre-RC complex; these kinases are the Dfb4-dependent kinase (DDK) and a cyclin-dependent kinase (CDK) and phosphorylate the MCM complex and CDC6, respectively (Jiang et al., 1999a; 1999b). There are a growing number of other factors involved in activation of the replication origin and these include Cdc45 and the four GINS proteins that, together with phosphorylated MCM2-7, form the CMG complex (Ilves et al., 2010; Im et al., 2009). The licenced origin is now activated and DNA

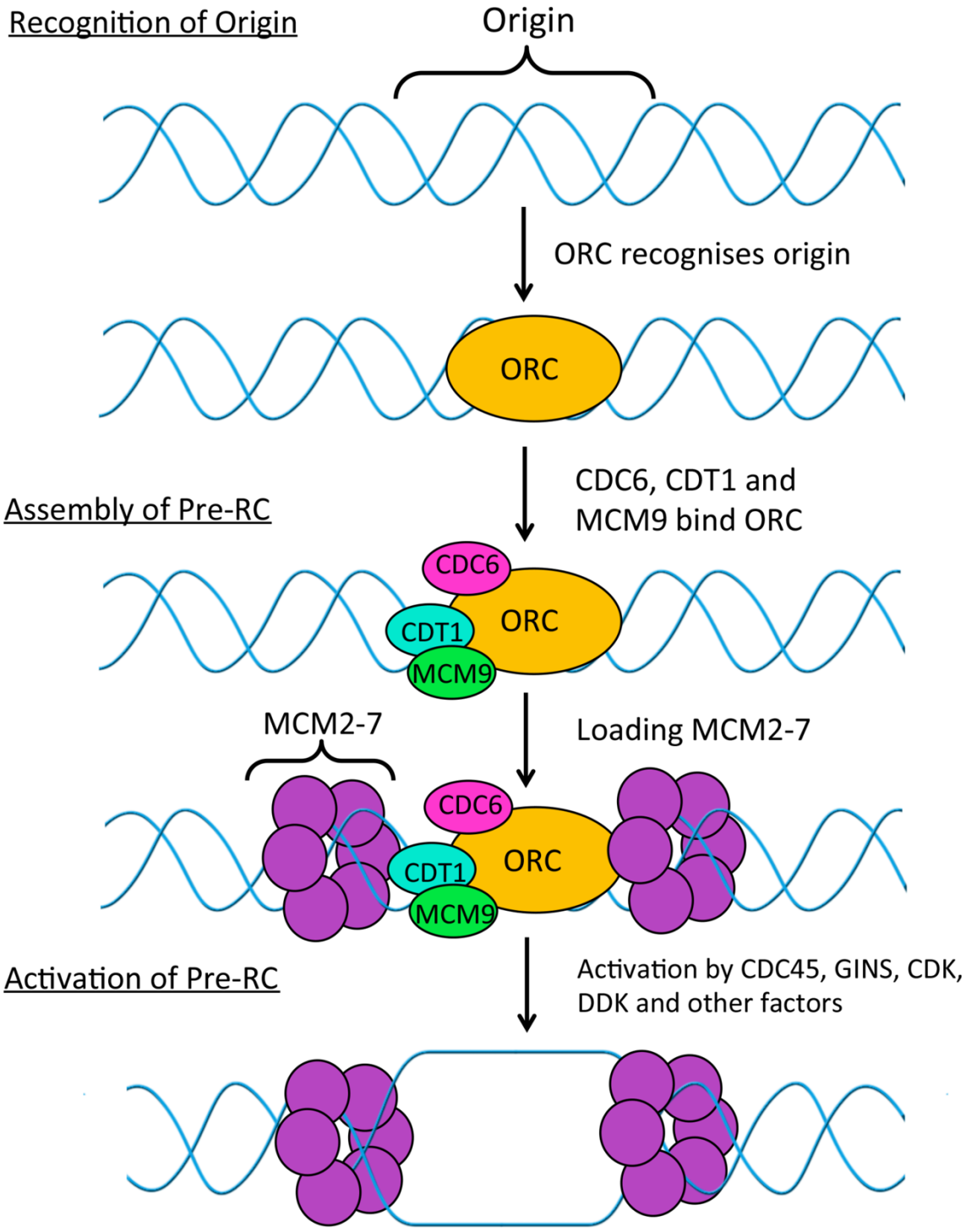


Figure 1.1. Assembly of the pre-replication complex (pre-RC) at DNA replication origins in eukaryotes.
The heterohexameric origin of replication (ORC) complex recognises a DNA replication origin. ORC recruits cell division cycle 6 (CDC6) and CDT1. CDT1 acts in concert with minichromosome maintenance protein 9 (MCM9) and, along with CDC6, loads the MCM2-7 helicase complex in an ATP-dependent manner to form the pre-RC complex. MCM2-7 forms a CMG complex with CDC45 and the four GINS and requires a number of other factors for pre-RC activation.

begins to be unwound by the helicase complex. Despite the CMG complex being loaded onto double stranded DNA, the CMG complex is remodelled once loaded and forms a single-stranded 3'-5' DNA translocase with a steric exclusion unwinding mechanism (Fu et al., 2011). The pre-RC complex is regulated by geminin (Lutzmann et al., 2006; Rao and R. T. Johnson, 1970). Geminin prevents rereplication by binding Cdt1 that is already undergoing licensing. Not all origins will lead to initiation of replication in every cell cycle and the question is therefore raised as how individual origins are selected. DNA combing studies have suggested that the firing of origins is stochastic and independent between cell cycles, with no genetic or epigenetic markers to prevent one origin from firing in the next round of replication (Patel et al., 2005).

1.2.2. Initiating Genome Replication

Once the pre-RC has been licenced and activated and the cell has progressed into S phase, genome replication can begin. The replisome forms at the bubble caused by the helicase unwinding by MCM2-7. Copying DNA occurs in two major steps, production of a short RNA primer and elongation using deoxyribonucleotides to produce a DNA copy. There are three main polymerases involved in replicating the bulk of DNA, these are DNA polymerases α, δ and ε; these polymerases all replicate DNA in a 5'-3' direction. The first, DNA polymerase α (pol α), has both DNA priming and polymerase activity and is the only polymerase of the three capable of initiating de novo primer synthesis (reviewed in (Burgers, 1998)). Historically, CDC45 has been the main candidate for recruitment of pol α to the replication complex. There are a number of supporting factors for this, such as the temporal coregulation of the two proteins *in vivo* as well as physical interactions between CDC45 and pol α in *Xenopus* and humans (Aparicio et al., 1999; Kukimoto et al., 2001; Mimura and Takisawa, 1998). However, a physical interaction was not observed in S. cerevisiae and an initiation defective mutant of CDC45 in S. pombe did not affect chromatin loading of pol α to the DNA, suggesting that CDC45 is not directly involved in pol α loading (Aparicio et al., 1999; Uchiyama et al., 2001). MCM10 is a protein that is required for the stimulation of phosphorylation of the MCM2-7 helicase by DDK and is another candidate for the recruitment factor of pol α (Lee, 2003). There is an increasing body of evidence suggesting that MCM10 is responsible for recruitment of pol α to the complex, including a number of studies suggesting its involvement in binding pol α and stabilising its association to chromatin (Chattopadhyay and Bielinsky, 2007; Ricke and Bielinsky, 2004; Warren et al., 2009; X. Yang et al., 2005; W. Zhu et al., 2007). Recently however, it has been revealed that a

trimer of Ctf4 may tether pol α to the CMG complex (A. C. Simon et al., 2014), both pol α and the Sld5 subunit of GINS contain Ctf4 binding motifs that bind Ctf4, this interaction is confirmed crystallographically and through size-exclusion chromatography but the interaction *in vivo* remains to be proven. Upon loading of pol α, it synthesises a short RNA primer of up to 15 nucleotides and extends from this about 20 nucleotides of DNA, from which other polymerases can extend (Conaway and Lehman, 1982a; 1982b; Kunkel, 2011).

Upon helical unwinding of the DNA, a protein known as replication protein A (RPA) is required to stop the DNA winding back on itself, forming secondary structures, or being degraded (reviewed in (Wold, 1997)). This protein binds pol α and was first identified through a Simian Virus 40 (SV40) replication system supplemented with mammalian cell extracts (Dornreiter et al., 1992; Fairman et al., 1988; Fairman and Stillman, 1988; Wold and Kelly, 1988). As SV40 requires only its large T antigen (TAg) from its own genome and all other replication components are derived from a mammalian host, this made it a prime candidate to study which proteins are required for the replication of its genome. These studies also identified the replication factor C (RFC), proliferating cell nuclear antigen (PCNA), topoisomerase I and topoisomerase II proteins as replication factors (Fairman et al., 1988; Prelich et al., 1987).

Following production of this DNA/RNA primer hybrid, the polymerase is switched to allow the replication complex to take over elongation. RFC is a heteropentameric complex that preferentially binds single stranded DNA and its main function is to load PCNA onto the DNA in an ATP-dependent manner (Jonsson et al., 1997; X. Li and Burgers, 1994). It has been demonstrated that the polymerase switching mechanism occurs in a two-step protein competition mechanism in which pol α is competed from RPA by RFC, which loads PCNA to the DNA, the replicative polymerase pol δ subsequently competes with RFC for RPA and PCNA (Yuzhakov et al., 1999). PCNA is a homotrimeric complex that encircles the DNA and is an example of a DNA clamp.

1.2.3. Elongation and Progression of the Replisome

The polymerase switching marks the switch to elongation. Elongation involves the highly processive, high fidelity replication of the bulk of the DNA by pols δ and ε. These two processive polymerases both have 3'-5' exonuclease (exo) activity to ensure fidelity (M. Simon et al., 1991; Tran et al., 1999). Whilst the SV40 studies had identified

a requirement for two polymerases, pol α and pol δ, studies in S. cerevisiae prior to this had found that a third DNA polymerase is required, pol ε (A. Morrison et al., 1990; Waga and Stillman, 1994). There is a topological problem in the replication of DNA at a fork that moves in a single direction, the two strands are antiparallel but all polymerases replicate DNA in a 5'-3' direction. The generally accepted model, initially provided by the Sugino lab and supported by the Kunkel lab, has been that pol α initiates replication prior to polymerase switching and the two pols, pol ε and pol δ, produce DNA on opposite strands, with pol ε producing a continuous strand of DNA opposite the leading strand and pol δ extending DNA primers in a discontinuous manner, producing disconnected strands of DNA opposite the lagging strand of approximately 200 nucleotides' length, known as an Okazaki fragment (Figure 1.2) (reviewed in (Kunkel and Burgers, 2008; Stillman, 2008)). There is a body of evidence that supports this model as it has been established that the exo of pol δ, but not of pol ε, corrects errors made by pol α, suggesting only pol δ has access to the primer termini on the lagging strand at the beginning of an Okazaki fragment (Pavlov et al., 2006). Exo mutations in pols δ and ε resulted in an increase the frequency of mutations on opposite DNA strands, suggesting that they do indeed act on opposite strands (Shcherbakova and Pavlov, 1996). Additionally, a number of studies into catalytic site mutations of pols δ and ε have shown these proteins to act on opposite strands in both fission yeast and budding yeast (Miyabe et al., 2011; Nick McElhinny et al., 2008; Pursell et al., 2007). There is emerging, however, a new distinct model based on recent evidence and discrepancies in evidence for the old model that, following elongation, as pol ε moves away from the origin of replication it dissociates from the DNA and is replaced by pol δ, which subsequently continues elongation (reviewed in (Pavlov and Shcherbakova, 2010)).

The Okazaki fragments produced through lagging strand synthesis must be joined to create a continuous strand of DNA opposite the lagging strand. As the polymerase reaches the RNA primer of the previously produced Okazaki fragment it peels away the primer, leaving a 5' overhang (or "flap") (Maga et al., 2001). This flap is cleaved by the flap endonuclease FEN1, which is stimulated through PCNA binding and can subsequently be ligated by DNA ligase I in an ATP-dependent manner (Barnes et al., 1990; Lehman, 1974; X. Li et al., 1995; Pascal et al., 2004).

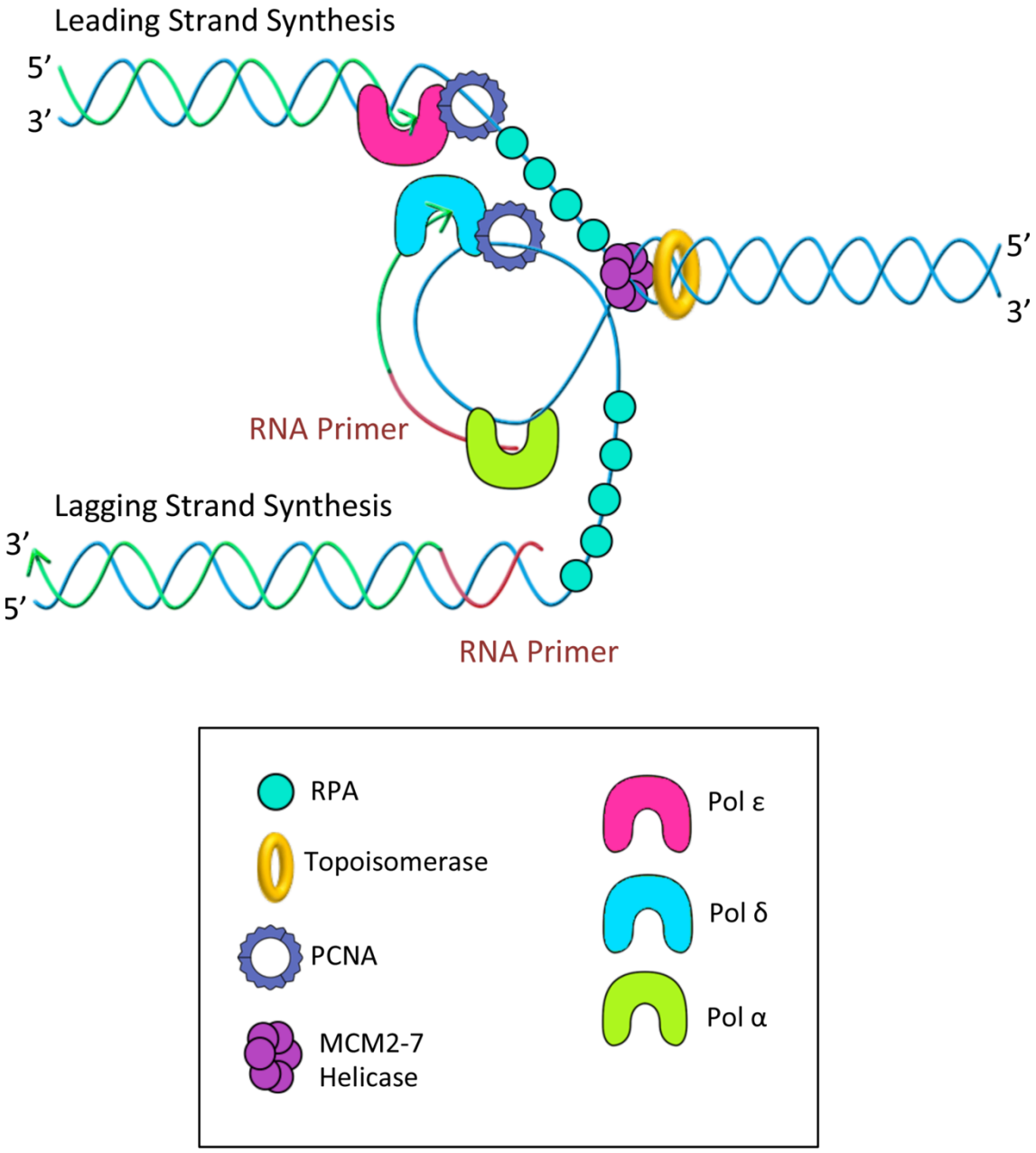


Figure 1.2. The eukaryotic replication fork.
The MCM2-7 helicase unwinds DNA at the replication fork. Torsional stress upstream of the replication fork is relieved by topoisomerases. Single-stranded DNA is bound by RPA to prevent formation of secondary structures and to protect from degradation. Leading and lagging strand replication both occur in a 5'-3' direction. The lagging strand is replicated in a discontinuous manner, with pol α producing a short RNA/DNA hybrid primer, before extension by pol δ. Leading strand replication occurs in a continuous manner with Pol ε extending primers produced by pol α. PCNA is expected to act as a processivity factor for pols δ and ε.

As a DNA helicase unwinds DNA and the polymerase replicates DNA, the DNA goes through a complete rotation every 10 nucleotides. As a result of positive supercoiling caused by this helicase unwinding, torsional stress upstream of the replisome is built up that needs to be released (Figure 1.3A). There are molecules that are specialised in the release of this torsional stress, known as topoisomerases. There are four types of DNA topoisomerases, type IA, IB, IIA and IIB and, of these, IB and IIA are the main candidates in involvement in replication fork progression (reviewed in (Champoux, 2001; J. C. Wang, 1996)). Type IB topoisomerases, such as topoisomerase I (Topo I), introduce a nick into the DNA and forms a covalent linkage with the DNA upstream of the nick through transesterification (Figure 1.3B). The two DNA segments can then rotate relative to one another prior to being reunited. Type IIA topoisomerases, such as topoisomerase II (Topo II) operate through an ATP-dependent manner to transport a DNA double helix through another by forming an enzyme-mediated gate in one of the double strands (reviewed in (J. C. Wang, 1998)). Topo I is likely to resolve positive supercoiling ahead of the fork, whilst also contributing to the prevention of interference between transcription and replication (Tuduri et al., 2009). Topo II, on the other hand, acts behind the fork to remove precatanae.

1.2.4. Genome Replication Termination

Replication forks initiate and elongate in a bidirectional manner and only come to a halt when fused with an adjacent replication fork. This fusion appears to occur randomly in a stretch of DNA with an average length of 5 kbp and contains fork pausing elements (Fachinetti et al., 2010; Greenfeder and Newlon, 1992a; J. Zhu et al., 1992). Unlike during transcription, termination of replication is not necessarily controlled by specific DNA sequences but primarily by where DNA replication is initiated and indeed, changing the site of initiation will change the site in which DNA replication terminates (Greenfeder and Newlon, 1992a). Within regions of termination, there are replication fork barriers that arrest one of the two forks in an orientation-dependent manner to allow the forks to fuse. There are a number of sites that slow DNA replication and could potentially result in replication fork fusion and these include, but are not limited to, sites of highly transcribed genes, centromeres and tRNA genes (Admire et al., 2006; Azvolinsky et al., 2009; Brewer and Fangman, 1988; Greenfeder and Newlon, 1992b). As the replication forks converge, topo I can no longer resolve supercoiling ahead of the fork and the fork must rely on topo II for relieving the torsional stress. Topo II can then remove the final catenations prior to segregation of the two DNA helices (Baxter

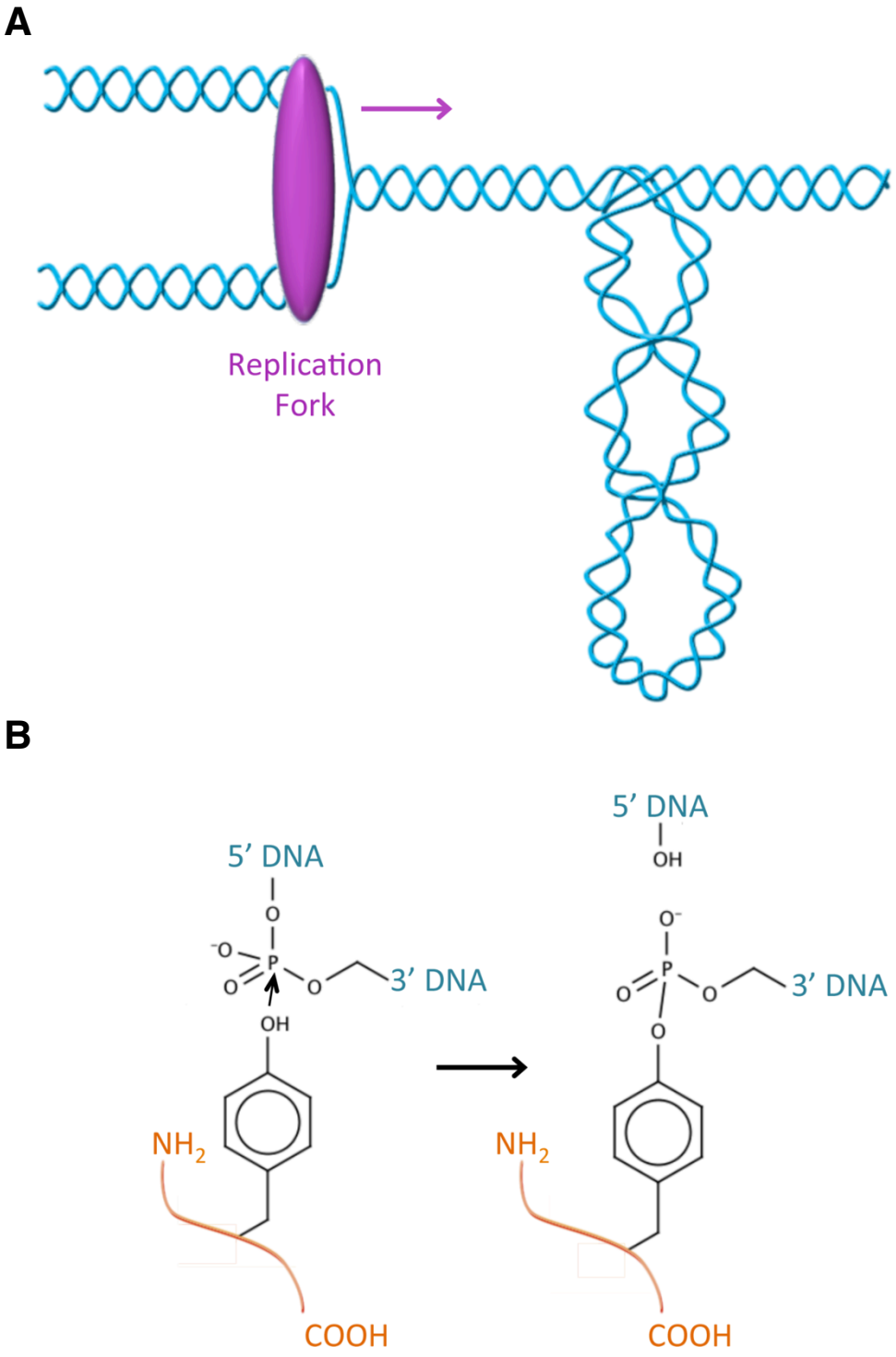


Figure 1.3: The action of topoisomerases.
(A) The helicase action of the replication fork introduces positive supercoils into the DNA upstream of the fork. The positive supercoiling results in torsional stress in the DNA. This stress is relieved by topoisomerases. **(B)** The mechanism of catalysis of type IB topoisomerases. The phosphate group of the DNA backbone undergoes nucleophilic attack by a tyrosyl group of the topoisomerase. This results in a covalent enzyme-DNA intermediate. An intact strand can be passed through the transient breakage to relieve the positive supercoiling.

and Diffley, 2008; Fachinetti et al., 2010; Sundin and Varshavsky, 1981). The mechanism by which this disassembly of the replisome occurs has previously been unknown until very recently through novel insights from a collaboration between the Labib and de Piccoli laboratories. The CMG complex (Cdc45-MCM-GINS) is transiently ubiquitylated during the final stages of chromosome replication in budding yeast (Maric et al., 2014). SCF^{Dia2} is an E3 ubiquitin ligase important for maintaining genome stability and it was found that there is no disassembly of the replisomes at the end of S phase without SCF^{Dia2}. This SCF^{Dia2} ligase is responsible for the ubiquitylation of the CMG complex and, upon examination of each of the subunits of the CMG complex, it was found that this ubiquitylation occurs only on MCM7(Maric et al., 2014). This ubiquitylated form of the CMG complex is only transient before the helicase complex is disassembled but this CMG-Ub is essential for disassembly**.** The transient nature of this ubiquitylation is Cdc48-dependent, and auxin-induced degradation of Cdc48 results in the persistence of this ubiquitylation (Maric et al., 2014).

1.2.5. Mitochondrial Genome Replication

Despite laboratory-cultivated vertebrate cells having the ability to grow anaerobically in the absence of mitochondrial DNA (mtDNA) multicellular organisms such as the simple nematode worm *Caenorhabditis elegans*, cannot survive without mtDNA (Bratic et al., 2009; Desjardins et al., 1985; King and Attardi, 1989). Mitochondrial DNA in protozoa exists as a closed circular DNA of around 15 kbp, encoding 22 tRNAs, 2 rRNAs and 13 mRNAs, which are translated into the proteins required for aerobic respiration (reviewed in (Wilson et al., 1985)). The replication mechanism in mitochondria is similar to nuclear DNA replication, with a number of corresponding proteins (Figure 1.4). The enzymes required for mtDNA replication include DNA polymerase γ, Twinkle DNA helicase, mitochondrial RNA polymerase (polRMT), DNA ligase III, RNase H1, mitochondrial transcription factor A (TFAM), single-stranded DNA binding protein (mtSSB) and topoisomerases (Cerritelli et al., 2003; Hance et al., 2005; Simsek et al., 2011; Spelbrink et al., 2001; Tyynismaa et al., 2004; Van Goethem et al., 2001). The nuclear helicases PIF1 and Dna2 have also been reported in the mitochondria (Duxin et al., 2009; Futami et al., 2007). Instead of using a specialised primase to generate primers, the mitochondrial transcription machinery produces short RNA transcripts at their origins of replication. Mitochondria have two distinct origins of replications on their two strands of DNA, the heavy strand (G-rich) and the light strand (C-rich). The heavy strand synthesis origin, known as O_H, is situated in a region of the mitochondrial

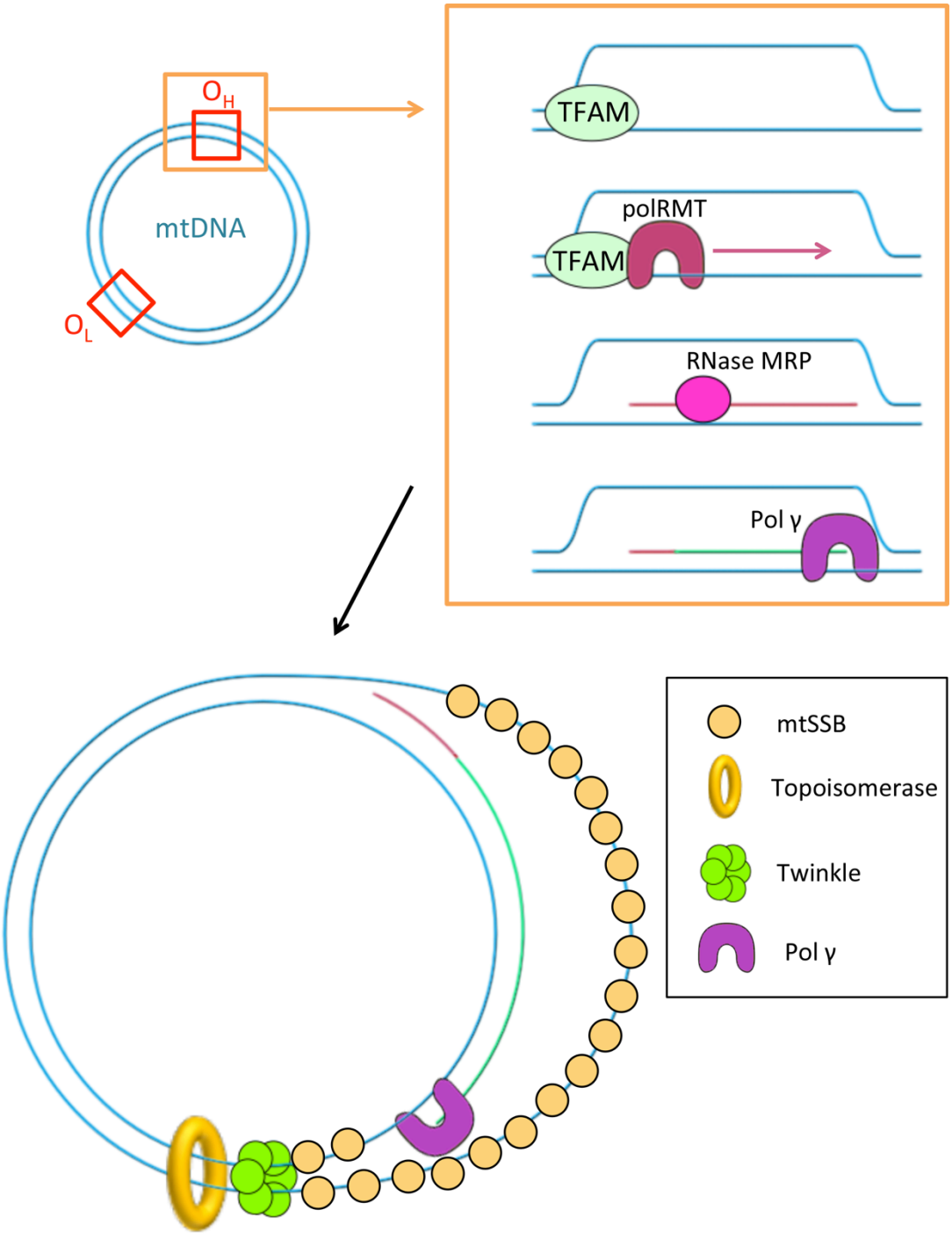


Figure 1.4. Replication of mitochondrial DNA.
Transcription Factor TFAM binds DNA at the origin of replication and unwinds part of the DNA. polRMT is then able to produce a strand of complementary RNA opposite the light chain of mtDNA. This RNA strand is processed by RNase MRP, the resulting primer is extended by pol γ. Twinkle helicase, along with topoisomerases can unwind the rest of the mtDNA whilst pol γ extends from the RNA primer around the closed circular DNA.

genome that contains all noncoding parts of mtDNA (Noncoding Region; NCR) (Pham et al., 2006). The light strand synthesis origin, known as O_L, is within a cluster of tRNA genes around two-thirds of the way around the genome relative to O_H and is not triggered until the heavy strand synthesis is around two-thirds complete (Chang and Clayton, 1985). TFAM belongs to a group of high-mobility group (HMG)-box family of proteins that characteristically bind and unwind DNA. TFAM binds the mtDNA upstream of the replication initiation site at O_H and permits the transcription of the unwound DNA (Dairaghi et al., 1995; Fisher et al., 1987). The next step is the production of an RNA primer from the origin of replication by polRMT (Chang and Clayton, 1985; Chang et al., 1985). The RNA primer is the same for transcription of mitochondrial genes and for replication of mtDNA but for replication it must be processed; this processing is carried out by a site-specific endoribonuclease known as RNase MRP (Topper and Clayton, 1990). The primer can then be extended by the A family polymerase DNA polymerase γ (pol γ). Pol γ is composed of two subunits, a 140kDa catalytic subunit and a 55kDa accessory subunit (Ropp and Copeland, 1996; Y. Wang et al., 1997). The RNA primer is recognised by the accessory subunit in a fashion similar to the recognition of mRNA by tRNAs and it is proposed that this enzyme has an additional role as a processivity factor, similar to PCNA but without the requirement for a clamp loader (Fan et al., 1999; Lim, 1999). The helicase used in mitochondria, Twinkle, was identified in patients exhibiting ophthalmoplegia and was noted for its similar sequence to the T7 primase/helicase, with its helicase domain exhibiting conserved organisation to other helicases (Z. Liu et al., 2008). Twinkle acts as a hexamer and unwinds DNA in a 5'-3' direction, it does this without a specialised protein to load it onto the closed circular mtDNA (Jemt et al., 2011; Korhonen et al., 2003). In combination with tetremeric mtSSB and pol γ, a processive replication machinery is formed that can synthesise long DNA molecules (see Figure 1.4) (Korhonen et al., 2004). The mechanism through which lagging strand synthesis occurs from O_L remains unknown and the primase that assists in the discontinuous synthesis of the lagging strand remains undetermined. The reason for the requirement of RNaseH1 and DNA ligase III has not been established, though given their activities elsewhere it is suggested that they are required for the removal of primers in the synthesis of the leading and/or lagging strand and the joining together of nicks resulting from the removal of primers and subsequent gap-filling.

1.3 DNA Polymerases

1.3.1 The Discovery of DNA Polymerases

Biochemistry in the era predating the discovery of DNA polymerases was predicated on work with the emphasis of energy production, as opposed to the metabolic processes of biosynthetic production. It was presumed that, since individual reactions could be reversed, entire pathways could be; this rationale seems reasonable when one considers that the production of glycogen from glycogen phosphorylase is a possibility – why not just reverse the action of exonucleases? After all, the idea that a mechanism by which an enzyme depended entirely on its substrate for instruction as a template was unprecedented in Biochemistry at the time. DNA was expected to be a primer for growth of a chain like glycogen was previously shown to be (Cori et al., 1943). During the 1950s, however, it became increasingly clear that the pathways for energy production and metabolic biosynthesis pathways were distinct, with biosynthetic pathways themselves having discrete synthesis and breakdown pathways.

The term “polymerase” was coined in 1958 upon the first purification and characterisation of an enzyme from *Escherichia coli* capable of performing nucleotidyl transfer of radioactive (^{14}C)-thymidine, from the form of dTTP, into DNA (Lehman et al., 1958). There are two basic features of the synthesis of DNA; the first is the nucleotidyl transfer, in which an active nucleoside triphosphate is transferred onto the growing strand of DNA following a nucleophilic attack on the triphosphate between the α and β carbons, removing two of the phosphate moieties, which provides the energy for the formation of a phosphodiester bond through a condensation reaction in an overall 5’-3’ direction (Kornberg, 1962). The second is the base pair matching to the template strand, this ensures fidelity of the genetic code between cells after genome replication; adenine is paired with thymine and cytosine with guanine, this accuracy is achieved through the large difference in enthalpy between matched base pairs and mismatched base pairs (Petruska et al., 1988). Polymerases must exhibit high processivity and fidelity if they are to replicate the genome rapidly and accurately enough for cells’ demands. Polymerases cannot, however, synthesise novel strands of DNA *de novo* as they cannot produce the initial short strands from which they can extend. This is carried out by specialised proteins, known as DNA primases (See section 1.7).

1.3.2. Structure and Function of DNA Polymerases

There are a large number of DNA polymerases that have been classified into seven different families, based on primary sequence and structural data: A, B, C, D, X, Y and

RT (reverse transcriptases) (Table 1.1) (Braithwaite and J. Ito, 1993; J Ito, 1991; Joyce and Steitz, 1994). Despite a huge variation in sequences between these families, nearly all polymerases share a common domain organisation and mechanism of action. The first polymerase structure to be resolved crystallographically was the C-terminal polymerase domain of *E. coli* pol I, known as the Klenow fragment (Klenow and Overgaard-Hansen, 1970; Ollis et al., 1985). The structure of the Klenow fragment in complex with DNA revealed further topological information about the polymerase (Beese et al., 1993). Following the initial structural studies of this Klenow fragment an analogy was drawn up of the polymerase resembling a right hand, with its subdomains referred to as the fingers, palm and thumb domains (Figure 1.5).

The domains form a crevice that has the palm subdomain at its base and is lined with the fingers and thumb subdomains. The substrate DNA is largely coordinated by amino acids in the thumb subdomain whilst the fingers subdomain is more important for the binding of the incoming nucleotide to be added to the nascent DNA strand. The polymerase active site, in which all the essential catalytic amino acids reside, is in the palm subdomain. This is an architecture shared across a diverse array of polymerases. Y family polymerases share the domain organisation of other polymerases but have a smaller fingers and thumb subdomains to give a more open, solvent-accessible active site (Reviewed in (Pata, 2010)). Additionally, they have a polymerase-associated domain (PAD; commonly also referred to as the 'little finger domain' in archaeal and bacterial systems).

The mechanism by which dNTPs are incorporated by polymerases is depicted in Figure 1.6. The first step in addition of nucleotides to a nascent DNA chain is the binding of the DNA substrate. There is a minimal change in the fingers and palm domain in the binding of DNA but there is a conformational change in a helix-loop-helix motif the thumb domain that closes in on the DNA (Figure 1.7; (Y. Li et al., 1998a; 1998b)). The dNTP is bound in the fingers domain, but distal (10-15 Å) from the active site, in the second step. The third step involves a large conformational change in the polymerase fingers subdomain (Figure 1.7; (Y. Li et al., 1998b)). There is a small 6° rotation in the core of the fingers subdomain, which is subsequently followed by a larger 40° rotation in the helices towards the tip of the enzyme, allowing the enzyme to close around the DNA and bring the incoming dNTP into contact with the catalytic centre. This conformation change step is vital for the discrimination of incorporation of

Family	Examples	Functions	Exonuclease Activity
A	T7 DNA polymerase, E. coli pol I, Taq pol I, Eukaryotic pols γ, ν, θ.	Chromosomal and mitochondrial DNA replication. Filling in gaps from the removal of nucleotides. Okazaki fragment processing.	5'-3' for the removal of Okazaki fragments. 3'-5' for proofreading.
B	Eukaryotic pols α, δ, ε, ζ, T4 DNA polymerase, E. coli pol II, Some viral and archaebacterial polymerases.	Chromosomal DNA replication. These enzymes are generally tethered by a processivity factor and stimulated by a single-stranded binding protein. Pol ζ used in DNA damage bypass.	3'-5' for proofreading.
C	Bacterial DnaE, E. coli pol III, Bacillus subtilis pol III.	Chromosomal replication. DNA damage bypass.	3'-5' for proofreading.
D	Euryachaeota polymerases including Pyroccocus furiosus DNA polymerase II.	Poorly characterised but suggested to be a replicative polymerase	3'-5' for proofreading.
X	Eukaryotic pols β, μ, λ, Yeast pol IV, African Swine Fever Virus pol X, Human Terminal Deoxynucleotide Transferase.	Involved in DNA damage processing including base excision repair and double strand break repair	-
Y	Eukaryotic pols η, κ, ι and REV1, E. coli DinB, UmuC.	DNA damage bypass. V(D)J	
RT (Reverse Transcriptase)	HIV RT, Moloney murine leukaemia virus RT, Hepatitis B RT, Telomerase	Reverse transcription of viral genomes. Telomerase has an innate RNA component to synthesise GC-rich repeats at telomeres	-

Table 1.1. The families of DNA polymerases.
An overview of the seven families of DNA polymerases describing examples of each family and their putative functions in vivo, as well as whether or not the family of polymerases generally have exonuclease activity.

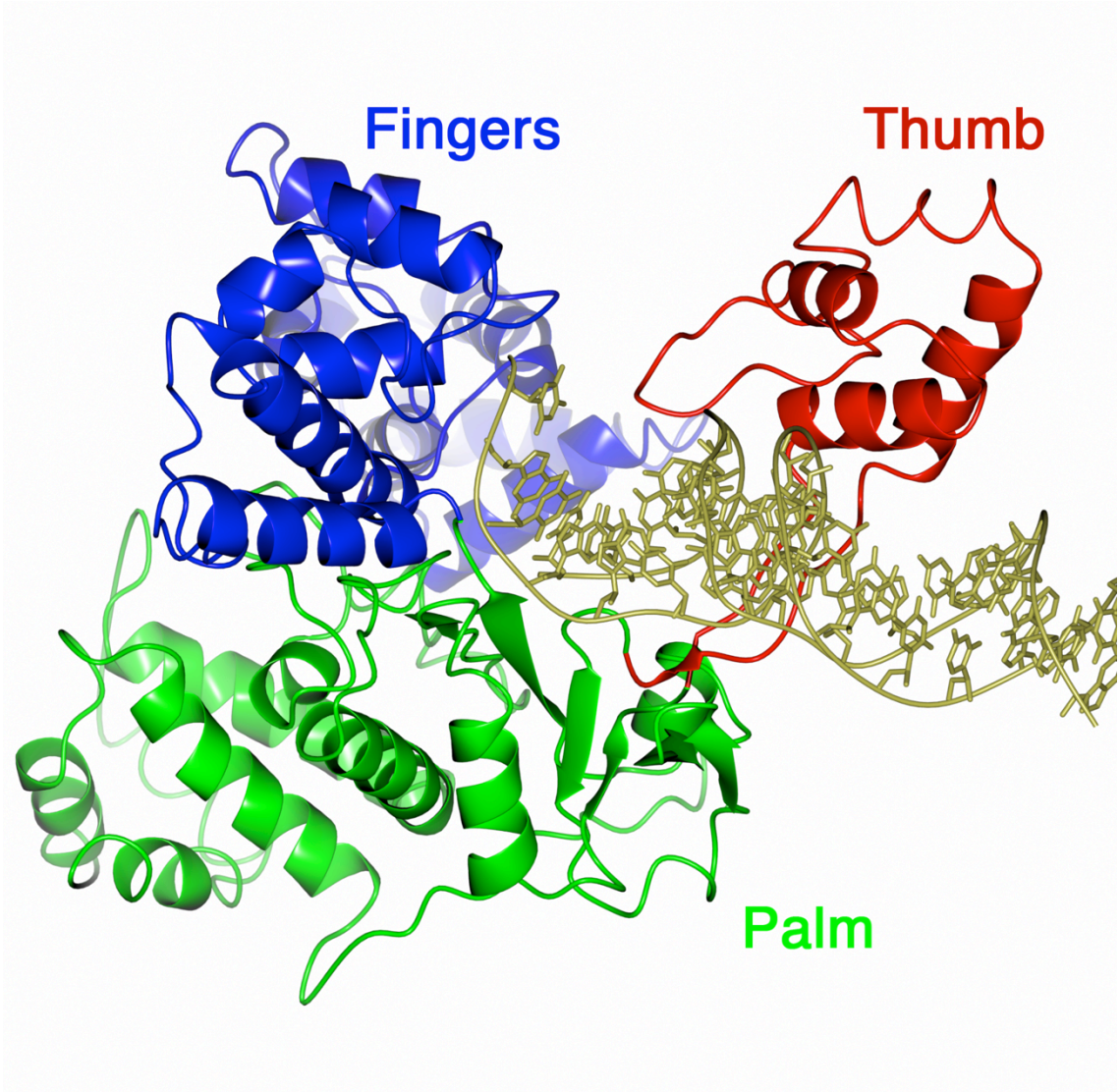


Figure 1.5. The structure of *Thermus aquaticus* (Taq) polymerase.
The structure of the Klenow fragment of Taq polymerase (PDB ID = 3E0D) resembles a right hand with its fingers (blue), thumb (red) and palm (green) subdomains through which the DNA is threaded. The template and nascent DNA strands are shown in gold.

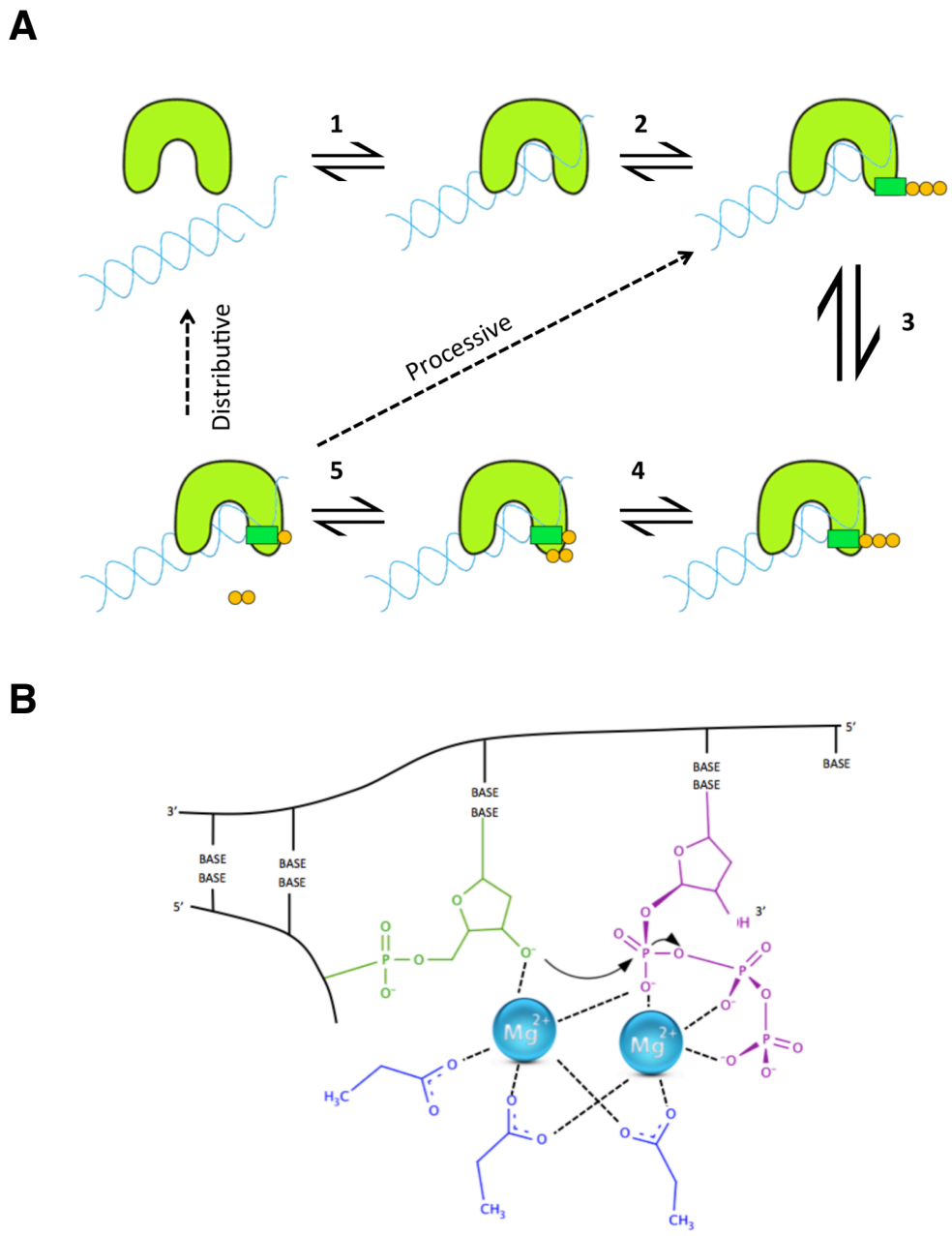


Figure 1.6. The Kinetic Pathway and Mechanism of Polymerase Catalysis.
(A) 1.) The nascent DNA chain binds to the polymerase. 2.) The binding of a nucleotide to the polymerase forms a polymerase-DNA-dNTP complex. 3.) The polymerase undergoes a conformational change to bring the nucleotide close to the active centre. 4.) The nascent DNA chain binds to the nucleotide and forms a new phosphodiester bond. The formation of a phosphodiester bond results in the release of a pyrophosphate ion. 5) The pyrophosphate ion is released from the enzyme ready for the next nucleotide addition. Distributive enzymes may release the DNA and begin at step 1, processive enzymes will remain bound to the DNA and continue from step 3. **(B)** A catalytic triad of amino acids (blue) coordinates two magnesium ions (light blue). These magnesium ions orientate the 5' base of the reaction (green) and the incoming dNTP (purple) to facilitate the nucleophilic attack on the α-phosphate, resulting in a pentavalent intermediate state. The reaction continues through an elimination reaction that removes the pyrophosphate from the α-phosphate.

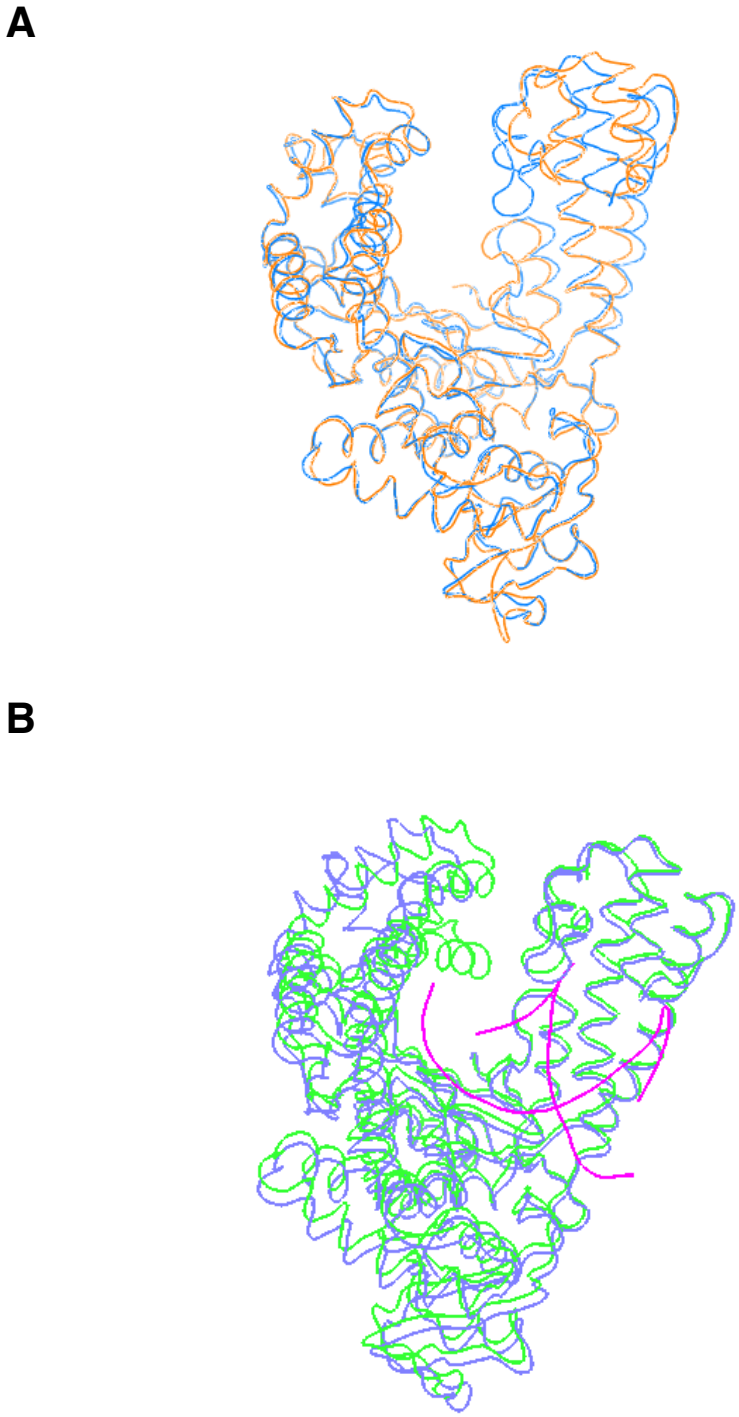


Figure 1.7. Conformational changes during Taq polymerase catalysis.
(A) Superposition of apo (orange; PDB ID = 5TKQ) and open DNA bound (blue; PDB ID = 2TKQ) Taq polymerase. Upon binding DNA, there is a small conformational change in the thumb domain from the apo Taq polymerase to the open conformation of DNA bound polymerase. The DNA is omitted from this figure for simplicity. **(B)** Superposition of open DNA bound (blue; PDB ID = 2TKQ) Taq polymerase and its closed (green; PDB ID = 4TKQ) conformation. There is a rotation in the fingers subdomain that closes the polymerase around the DNA (shown in magenta) to orient the DNA for catalysis.

dNTPs, as well as the phosphoryl transfer. The protein forms a close-fitting structure around the nascent base pair, providing a steric hindrance to the incorporation of incorrect base pairs (Y. Li et al., 1998a; Y. Li and Waksman, 2001). Due to the large conformational change in the DNA, it was initially presumed that this was the rate limiting step but a FRET analysis of this subdomain determined it to be significantly quicker than the kinetically determined rate-limiting step (Rothwell et al., 2005) and this rate-limiting step is now generally accepted to occur in the closed state.

The phosphoryl transfer reaction can then occur. Despite the protein scaffolds around the catalytic centres being divergent, the structures of polymerases have long been known to share a congruence in the coordination of two metal ions by acidic amino acid residues that are presumed to stabilise a pentavalent transition state caused by a nucleophilic attack in the polymerisation of oligonucleotide chains (see Figure 1.6B; (Steitz, 1999; Steitz et al., 1994)). Two divalent cations are held by three essential acidic amino acids 3.8Å apart. In some polymerases, all three amino acids are aspartates but in others two are aspartate residues and one is a glutamate; only the metal ions are directly required for catalysis with the acidic residues providing a structural framework in which to coordinate the ions. Whilst metal coordination differs between structures the metal ions are positioned identically relative to the primer terminus and α-phosphate of the incoming deoxynucleotide triphosphate (dNTP). This reaction is then followed by a further conformational change to allow the release of the pyrophosphate. Following the release of pyrophosphate, the next nucleotide can then be added to the nascent DNA chain or, if the polymerase is a distributive enzyme, the chain can be released (Figure 1.6A).

1.4 DNA Damage

DNA is not an inert molecule, rather the opposite, and it is susceptible to damage from the environment, whether this is the immediate endogenous environment or the wider exogenous environment. The integrity and stability of DNA are paramount for viability and genomic consistency and in Section 1.5 I will discuss the ways in which cells deal with damaged DNA. In this section we take a look at the insults that cause DNA damage, There are an estimated 50,000-100,000 chemical DNA modifications predicted to occur in a single cell every day (Hübscher and Maga, 2011). This section will discuss the most common causes of DNA damage.

1.4.1. Ionising radiation and its effects on DNA

Ionising radiation is radiation with enough kinetic energy to cause the release of an electron from an atom, thereby ionising it. Ionising radiation was the first external agent shown to be able to cause mutation in living organisms in *Drosophila melanogaster* fruit flies (Muller, 1927). Ionising radiation has the potential to act both directly and indirectly to alter the structure of DNA. Ionising radiation with high linear energy transfer, such as α particles, neutrons or β particles, can act directly on DNA. Ionising radiation such as X-rays can also interact with intracellular molecules including oxygen and water to produce reactive oxygen species, such as hydroxyl radicals, that can subsequently damage DNA (see section 1.4.3.).

Ionising radiation causes a wide spectrum of DNA lesions and these include damages to nucleobases, double-stranded DNA breaks (DSBs) and single-stranded DNA breaks (SSBs). SSBs are caused by the reaction of hydrogen atoms in the deoxyribose moiety of DNA and are readily repaired by ligation (Dean et al., 1969; Vanankeren et al., 1988). They are far more prevalent than DSBs (about 1000 $cell^{-1}$ Gy^{-1} compared to 25-40 $cell^{-1}$ Gy^{-1} for DSBs), which are more lethal to the cell as simple religation is not an option (Olive, 1998).

1.4.2. Ultraviolet Light and its Effects on DNA

Ultraviolet (UV) irradiation, mainly from solar sources, has a mutagenic effect on cells. At wavelengths approaching the maximum absorbance of DNA (i.e. 260 nm) adjacent pyrimidines can form bulky covalent linkages (Figure 1.8) (Friedberg et al., 1995; Rastogi et al., 2010). 70-80% of these covalent lesions are cyclobutane pyrimidine dimers (CPDs) and around 20-30% are pyrimidine (6-4) pyrimidone photoproducts ((6-4)PPs). The generation of 6-4 photoproducts is the result of an oxetane intermediate transferring a hydroxyl group at the carbon at the 4' position of the 3'-nucleobase to the carbon in position 6' of the 5' nucleobase. This product can then be converted to its Dewar valence isomer by irradiation of lower energy UV light with wavelengths of approximately 300-350 nm (Taylor, 1990; Taylor and Cohrs, 1987).

CPDs bend the helical axis by 30° and unwind the DNA by approximately 9°; the bases remain stacked enough for CPDs to continue to form Watson-Crick base pairs, albeit with one hydrogen bond lost between the 5'-thymine and adenine (Park et al., 2002). (6-4)PPs result in a much higher distortion to the DNA and the two adjacent bases are

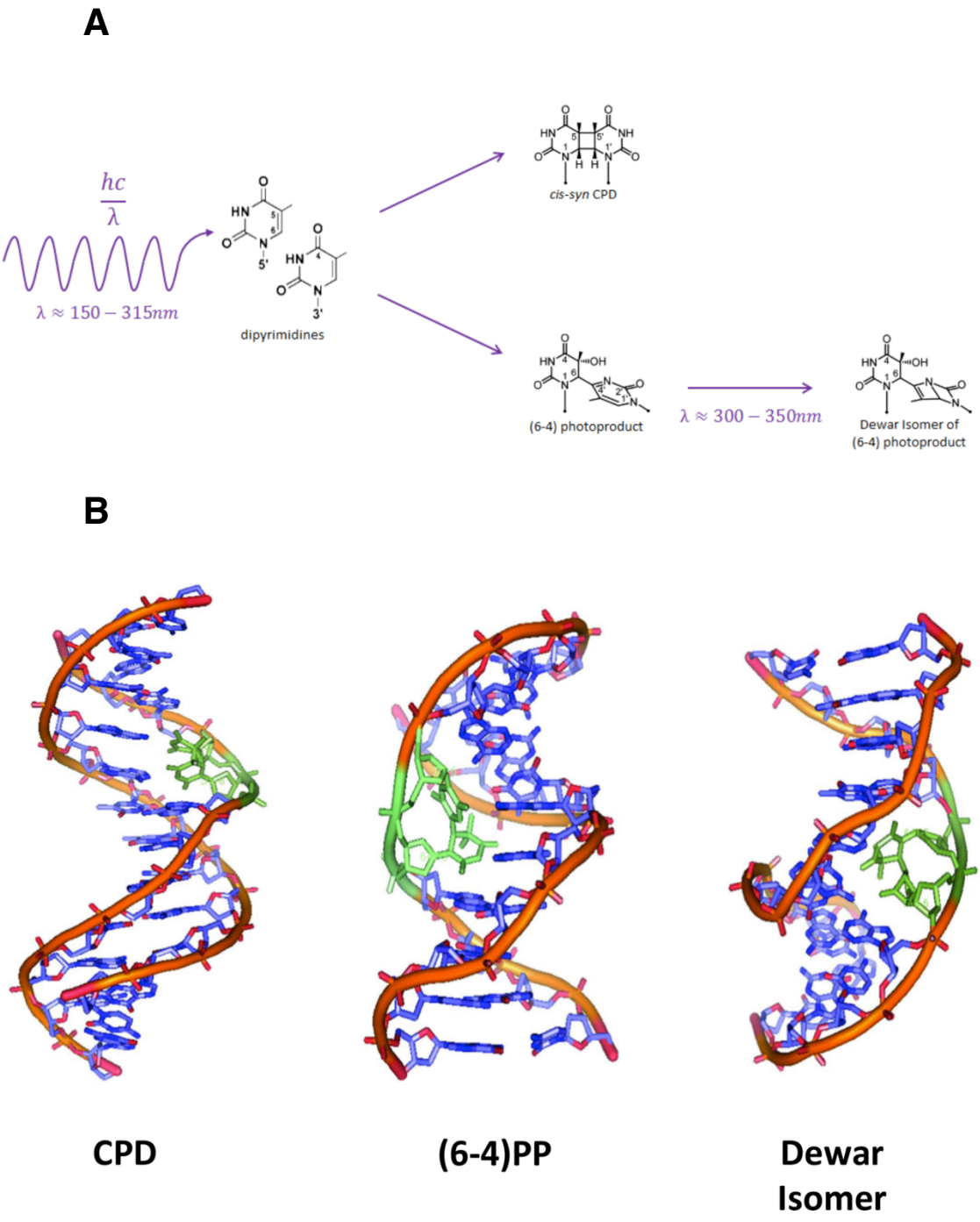


Figure 1.8. Formation and structure of different types of DNA lesions caused by UV irradiation of two adjacent (thymine) pyrimidines.
(A) Around 70-80% of covalent lesions caused by UV light are cyclobutane pyrimidine dimers (CPDs) and 20-30% of lesions are thymine-thymine pyrimidine (6-4) pyrimidone dimers ((6-4)PPs), which can subsequently be converted into the Dewar valence isomer. hc/λ is the energy of a photon, where λ represents the wavelength of the electromagnetic radiation, h is the Planck constant and c is the speed of the electromagnetic radiation. **(B)** Structures of DNA duplexes showing the presence of UV lesions (in green) of CPD, (6-4)PP, and (6-4)PP Dewar isomer. (B) adapted from Rastogi et al., 2010.

found at approximately right angles to one another. The helical bending at these lesions is estimated to be closer to 44° and phosphorus atoms in the paired bases are found within a distance of 6.13 Å as opposed to the 7.11 Å average at other base pairing steps.

1.4.3. Oxidative Damage to DNA

Guanine moieties in DNA have a high oxidative potential in cells (they are the nucleobases with the lowest E_{ox}) and readily oxidise to 8-oxo-2'-deoxyguanosine (8-oxo-G). A number of oxidising agents have been shown to induce 8-oxo-G formation (See Figure 1.9) (Halliwell and Aruoma, 1991). The fibres of asbestos cause their toxicity through the production of reactive oxygen species from hydrogen peroxide (Mahmood et al., 1994). Reactive radical species that induce the formation of 8-oxo-G within both isolated and cellular DNA are also produced by exposure to ionising radiation; this is predominantly carried out by a singlet oxygen species as the reactive intermediate (Devasagayam et al., 1991; Dizdaroglu, 1985).

The structure of the DNA double helix is such that a π-stacked array is formed by the heterocyclic base pairs, allowing the migration of charge across large distances; it has been revealed that oxidation of guanine to 8-oxo-G can occur up to 37Å away from the site of the reactive species, thereby increasing the likelihood of formation (Hall et al., 1996). Two adjacent guanine moieties are preferred sites for oxidative reaction to occur, with the 5' guanine being the more reactive (Prat et al., 1998). Quantum mechanical calculations of hole (areas of localisation of DNA radical cations) spatial distribution in ionised oligomers containing the sequence 5'-U/T-G-G-U/T-3' have approximately equal radical cation density on the guanines when the sequence is 5'-TGGT-3' and 5'-UGGU-3'. However, oxidation of 5'-TGGT-3' by electromagnetic irradiation gives a ratio of 5'-G oxidation to 3'-G oxidation of 6.1, while this ratio for 5'-UGGU-3' is just 3.4 (Prat et al., 1998). The methyl group of thymine, therefore, exerts a significant steric hindrance of the formation of 8-oxo-G on the 3' guanine. As a result, it can be deduced that both steric and electronic factors affect the rate of reaction (referred to as k_{trap}, as it is the rate at which electrons are irreversibly trapped at a site of oxidation as opposed to hopping between bases). 8-oxo-G is mutagenic to cells because, whilst 8-oxo-G can be overcome by the replicative polymerases DNA pols α, δ and ε, the replicative enzymes all most commonly insert an incorrect adenine opposite an 8-oxo-G as opposed to the correct cytosine as a result of Hoogsteen base

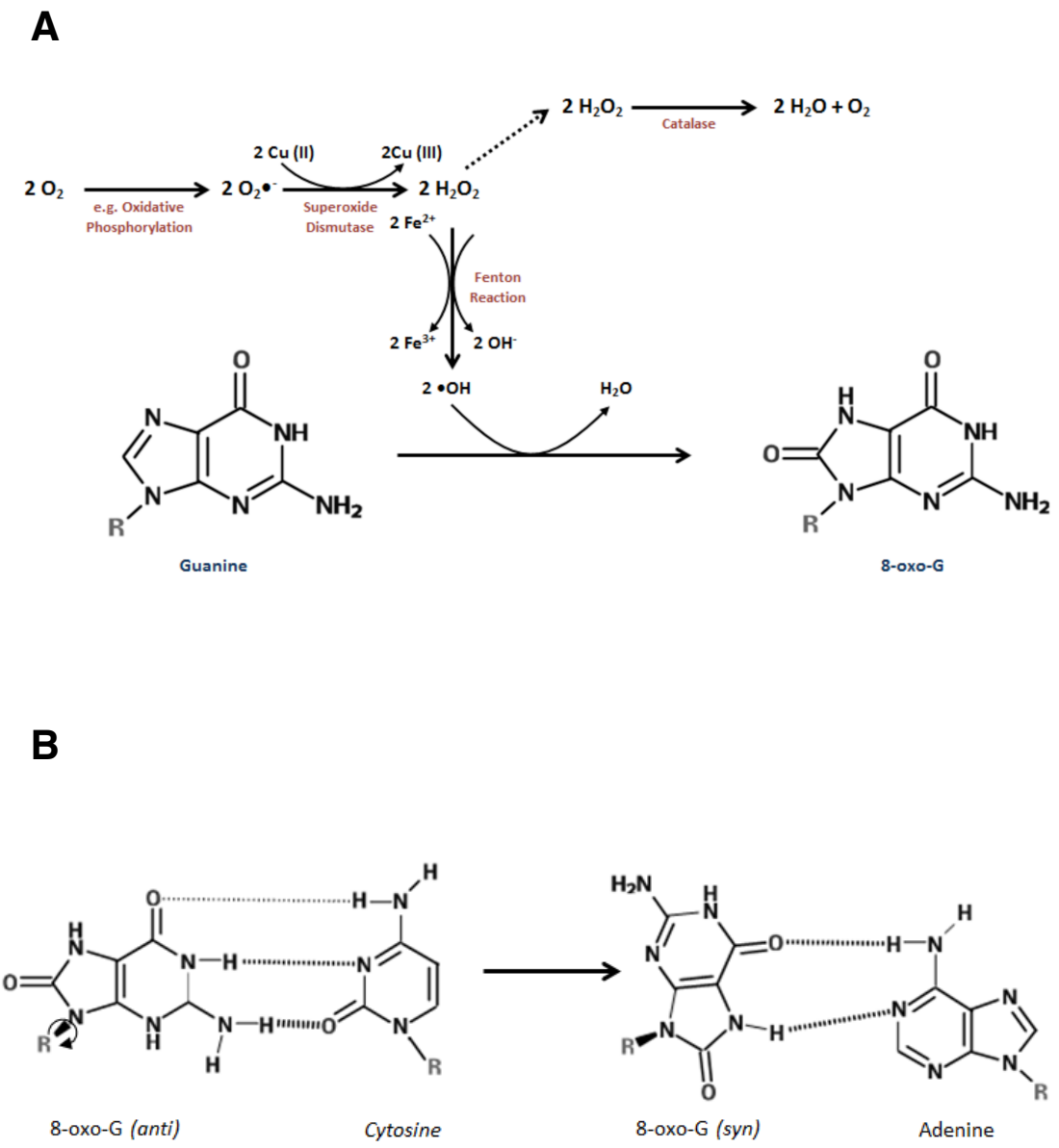


Figure 1.9. Oxidation of guanine and its subsequent mutagenicity.
(A) Formation of 8-oxoguanine (8-oxo-G) from guanine as a result of superoxide molecules and other reactive oxygen species in the cell. Superoxide radicals can be found in the cell as a result of oxidative phosphorylation in the mitochondria or through the action of ionising radiation. These radicals can be metabolised to hydrogen peroxide, which can be used to oxidise iron molecules and results in the release of hydroxide ions and hydroxyl radicals. These hydroxyl radicals react with guanine to produce 8-oxoguanine. **(B)** 8-oxoguanine in its *anti-* conformation can form Watson-Crick base pairs with cytosine, as guanine does. However, in its *syn-* conformation, it forms Hoogsteen base pairs with adenine. Replicative polymerases constrain 8-oxoguanine to its *syn-* conformation and incorporate an incorrect adenine opposite this damaged base.

pairing, thus generates a GC → TA transversion following replication (Figure 1.9) (van Loon et al., 2010).

Despite the fact that adenine has a lower E_{ox} than thymine, if there is no suitable guanine moiety available, oxidation occurs at two adjacent thymine moieties with the formation of thymine glycol (Tg), among other products, in a process initiated by the hydration of the thymine radical cation carbon double bond (Ghosh et al., 2008). This is because the activation free energy for the thymine radical cation is lower than that for reaction at the adenine radical cation; the relative yields of products is determined by the difference in free energy of the transition states ($\delta\Delta^{\ddagger}G$), not by the relative abundance of the intermediates, which would indeed be determined solely by E_{ox} (Curtin-Hammett principle).

1.4.4. Chemical Agents and Their Effects on DNA

There are a huge number of chemicals that can induce DNA damage, including oxidising agents (see Section 1.4.3.), alkylating agents and intercalators. Bifunctional alkylating agents such as nitrogen mustards and cis-platin and their related compounds have been used in cancer chemotherapy for a long time; these molecules crosslink DNA (Rink et al., 1993; Rosenberg et al., 1969).

The type of adduct caused by alkylating agents depends on their nucleophilicity; those with a high nucleophilicity, such as MMS, will undergo bimolecular (S_N2) substitutions, whereas those with a low nuclepohilicity, such as N-ethyl-N-nitrosurea (ENU), will undergo unimolecular (S_N1) substitutions (Hoffmann, 1980). S_N2 substitutions generally result in adducts at the nitrogens at position 7 of guanine and position 3 of adenine. The high nucleophilic reactivity of the nitrogen at position 7 of guanine and position 3 of adenine results in the formation of methylated adducts N7-methylguanine (7meG) and N3-methyladenine (3meA) respectively as induced by alkylating agents. 7meG is the major methylated adduct (60-80% of methylated lesions) but is not itself mutagenic or cytotoxic; however, it is prone to the formation of apurinic/apyrimidinic (AP) sites by depurination (See Section 1.4.5.). Conversely, 3meA, which accounts for around 20% of methylated lesions, can block the DNA polymerase machinery and is thus highly cytotoxic (Drabløs et al., 2004). S_N1 substitutions are less common; they are also likely to cause adducts at the nitrogen of position 7 of guanine but can also alkylate the

oxygen at position 6 of guanine (Hoffmann, 1980); O6-methylguanine mispair with thymine causing A→T transition mutations.

Another common covalent modifier of DNA is Benzo[a]pyrene, which itself is a procarcinogen found in cigarette smoke and pollution and is sequentially oxidised by cytochrome P450 to benzo[a]pyrenediolepoxide (BPDE) (Sims et al., 1974). BPDE is carcinogenic and can covalently bind DNA (see Figure 1.10) as well as intercalating within DNA to induce mutation (Geacintov et al., 1984).

Intercalating agents do not cause any direct covalent modifications to DNA, but rather interfere with base stacking and can induce point mutations. These aromatic compounds unwind DNA and fit between the base pairs. Anthracyclines are a class of intercalators that are currently used in chemotherapeutic treatment and include daunorubicin and doxorubicin (reviewed in (Minotti et al., 2004)). These molecules consist of a common sugar moiety (daunosamine) and a tetracyclic aglycone group.

1.4.5. Endogenous DNA Damage

In addition to naturally occurring oxidative damage and alkylation, there are is a huge risk from endogenous sources to the integrity of DNA. Hazards to DNA stability include spontaneous depurination, spontaneous deamination, DNA replication errors and chromosomal rearrangements.

Depurination of nucleotides by hydrolysis occurs a predicted 9000 times per cell per day (Nakamura et al., 1998). It is a consequence of the removal of the base at the glycosyl bond and results in an AP site. AP sites can inhibit DNA replication but also cause base substitution mutations (Loeb and Preston, 1986). As AP sites are found as an intermediate in base excision repair (BER; see section 1.5.2.) they are often repaired by this machinery.

Spontaneous deamination of cytosine results in uracil bases (Figure 1.11) (Duncan and Miller, 1980). Uracil is found naturally in RNA in cells and base pairs to adenine; as a consequence of this adenine base pairing, cytosine deamination results in G→A transition mutation. Additionally, uracils are removed from DNA by the enzyme uracil DNA glycosylase and this removal leads to an AP site.

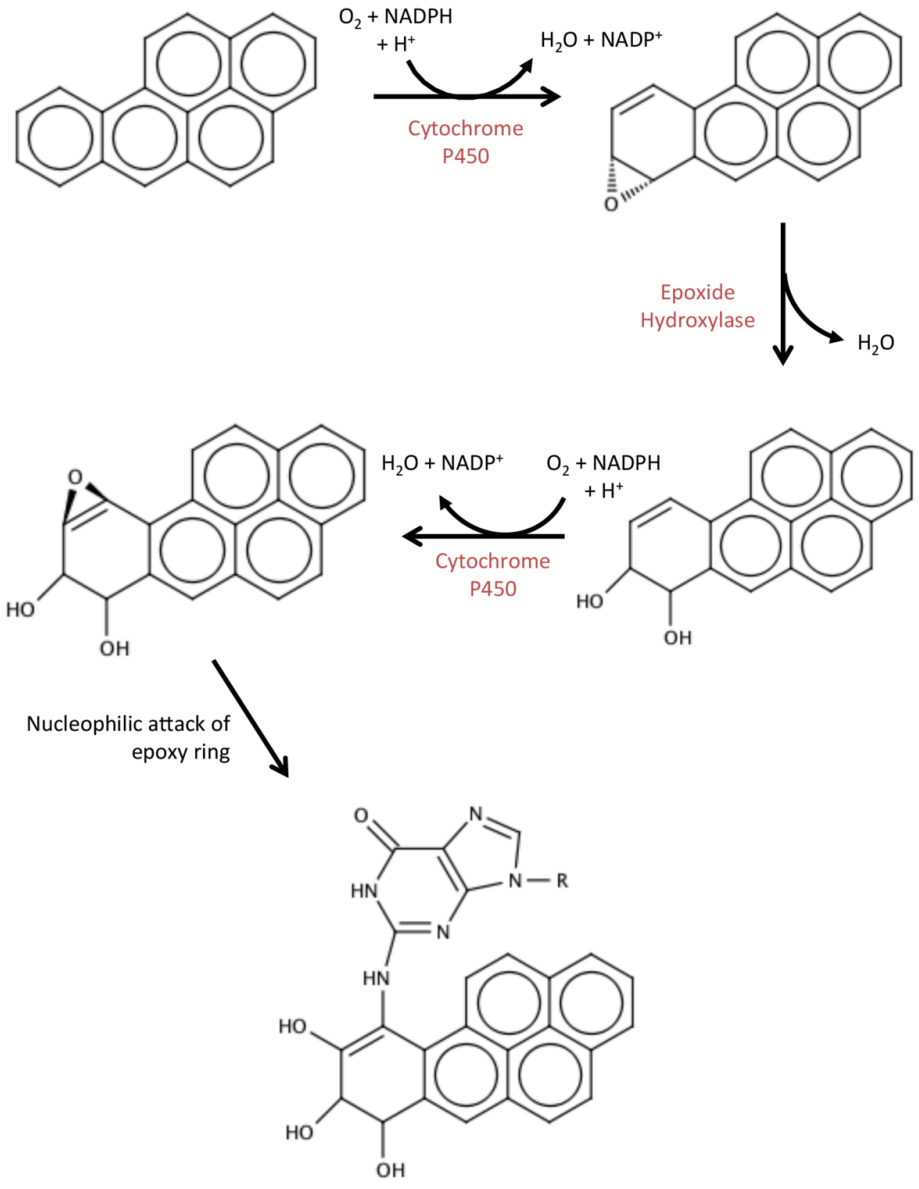
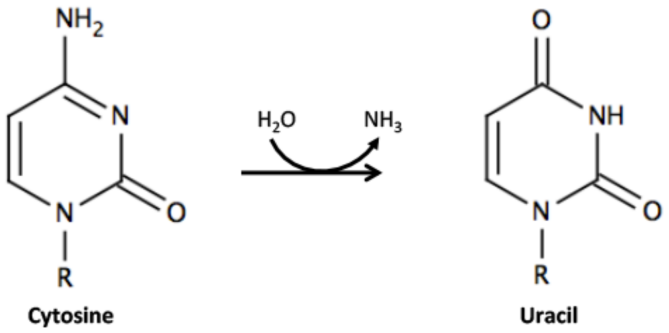


Figure 1.10. Covalent modification of guanine by benzo[a]pyrene.
Benzo[a]pyrene is oxidised by cytochrome P450 to produce an epoxy intermediate. This epoxy intermediate is then hydrolysed by epoxy hydrolase to form a diol. This diol is further oxidised by cytochrome P450 to produce Benzo[a]pyrene diolepoxide (BPDE). The amine group of guanine then attacks the epoxide ring to form a mutagenic bulky adduct.

A



B

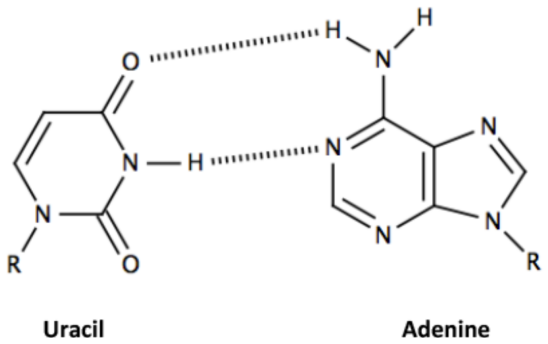


Figure 1.11. Deamination of cytosine to uracil and its subsequent mutagenicity. **(A)** Cytosine undergoes spontaneous hydrolysis and is deaminated to a uracil base. **(B)** If uracil bases persist into replication they will base pair with adenine resulting in transition mutations.

Whilst estimates of mutation rates are as low as 1.2 x 10^{-10} mutations/base pair/cell, a lot of this is attributable to post replication repair (Loeb, 1991). The main replicative enzymes, pol δ and pol ε, have inherent replication errors of 4.5 x 10^{-6} and 2 x 10^{-5} mutations/base pair respectively (Schmitt et al., 2009; Shcherbakova and Pavlov, 1996). Occasionally, polymerases will also insert or delete bases as a result of substrate misalignment. Breakages in the DNA that are subsequently rejoined in an alternative arrangement can also lead to chromosomal rearrangements. One famous example includes the fusion gene created through a rearrangement of the BCR gene on chromosome 22 to chromosome 9 adjacent to the Abl1 gene to produce an elongated chromosome 9 known as the Philadelphia chromosome. This BCR-ABL gene fusion product is a highly oncogenic tyrosine kinase and is responsible for chronic myelogenous leukaemia and some forms of acute lymphoblastic leukaemia (Deininger et al., 2000).

1.5 DNA Damage Repair

Damaged DNA must be repaired to ensure cell viability. Cells that don't repair their damage can become senescent, apoptotic or malignant (Bertram, 2000; Collado et al., 2005). Defects in DNA repair pathways are the cause of a number of hereditary diseases, including ataxia-telangiectasia (AT) and xeroderma pigmentosum (XP). AT patients are unable to repair oxidative damage in the cerebellum, resulting in a degenerative motor condition and have a weakened immune system (Savitsky et al., 1995). XP patients have defects in their UV repair pathway and show an increased sensitivity to UV light (Lehmann, 2003). The different types of DNA repair pathways will be discussed in this section, with an overview shown in Figure 1.12.

1.5.1. Direct DNA Damage Reversal

Methylation of guanine and damage by UV light are both known to be reversed directly by specialised enzymes known as O6-methylguanine-DNA-methyltransferase (MGMT) and photolyase respectively. Photorepair of UV damage caused by UV irradiation was first demonstrated in 1949; it was shown that cells rendered non-viable by UV irradiation that were then exposed to higher wavelength light (visible light) for longer periods of time showed a greater recovery to viability than those that were kept in the dark (Kelner, 1949). This photoreactivation is a direct, high fidelity and energetically favourable method of reversing pyrimidine linkages to their monomeric form; photoreactivation is catalysed by an enzyme known as photolyase (Minato and Werbin,

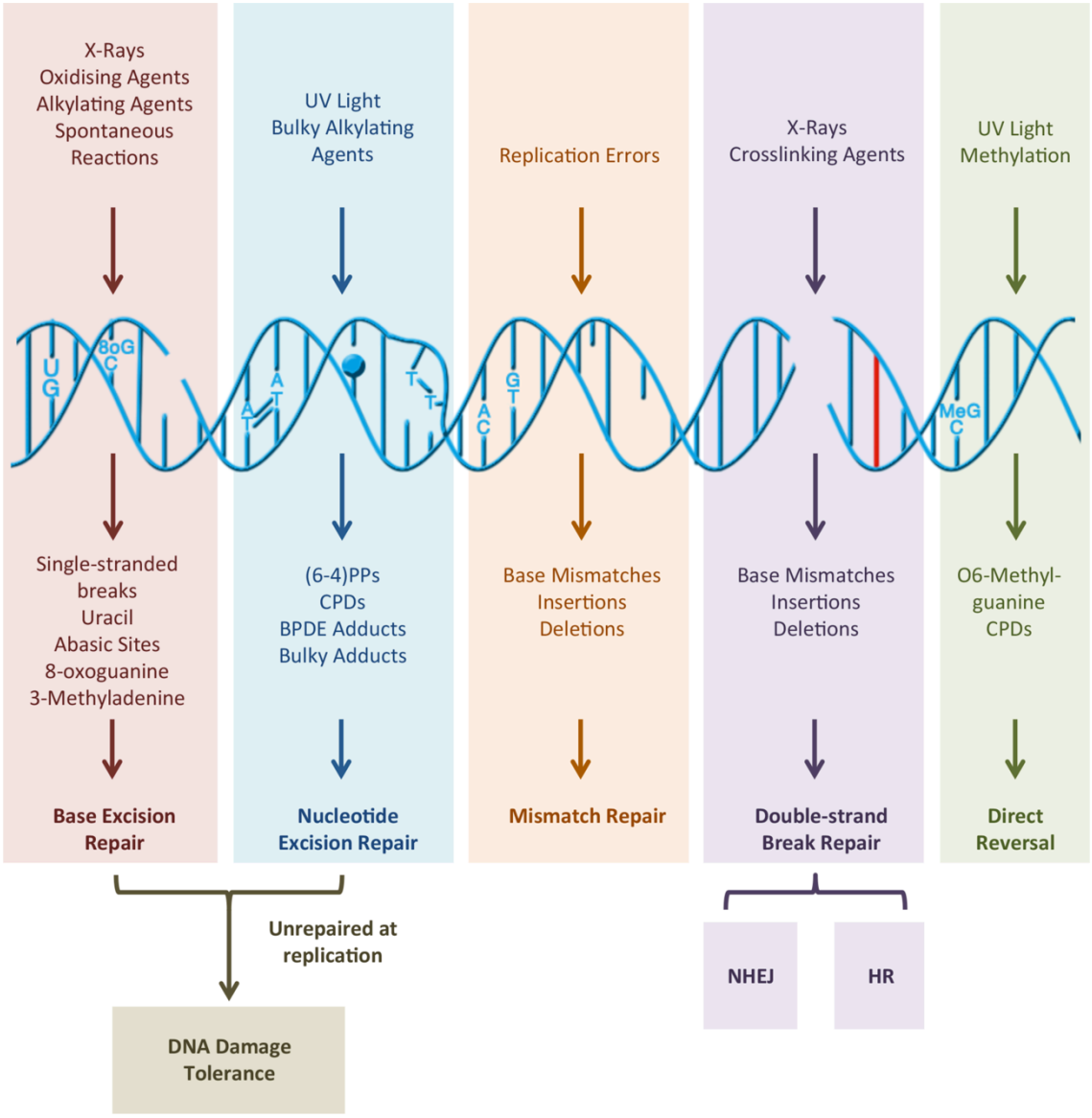


Figure 1.12. Cellular repair mechanisms of the damage caused to DNA.
DNA is damaged through a wide variety of exogenous and endogenous chemicals and high-energy ionising radiation. The DNA becomes damaged in a number of different ways and these different types of damage are removed by specialised repair mechanisms including base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), double-strand break repair and direct reversal. Non-homologous end-joining (NHEJ) and homologous recombination (HR) are two mechanisms through which double strand breaks are repaired. DNA damage that is unrepaired prior to genome replication undergoes DNA damage tolerance at the replication fork.

1972). These proteins are monomeric, and range between 53 and 66kDa, depending on the organism (G. B. Sancar, 1990). Photolyases can be found in fish, reptiles and marsupials, but appear to be absent in humans and other placental mammals. There are two related genes in humans (CRY genes) but encode blue light photoreceptors involved in setting circadian rhythms, not photoreactivation (Todo et al., 1996).

Photolyases depend on the cofactor flavin adenine dinucleotide (FAD). The UV lesion is flipped out of the DNA enters a cleft in the protein between two helix bundles wthin which FAD is held in a 'U-shaped' conformation (Park et al., 1995). FAD causes a reductive cleavage, leading to the breaking of CPDs and (6-4)PPs (A. Sancar, 2003). FAD is a redox catalyst with the potential for one- and two- electron transfer reactions; it is a ubiquitous molecule and is involved in a number of metabolic processes. In its fully reduced form (i.e. catalytically active), FAD absorbs light in the blue (visible light) and near-UV ranges of visible light (G. B. Sancar, 1990). Additionally, a second cofactor with a large extinction coefficient acts as a photoantenna that absorbs light and, through resonant energy transfer to FAD, improves the reaction efficiency. Organisms that do not possess photoreactivation-capable enzymes must find an alternative method of tolerating UV damage.

DNA methylated at position O6 (See Section 1.4.4.) can be repaired by MGMT. MGMT stoichiometrically repairs DNA adducts and self-inactivates, it is therefore a very energy rich mechanism of repairing methylation. MGMT operates by the transfer of the methyl group of the guanine-associated adduct to a cysteine residue in the catalytic cleft of the enzyme, thus restoring the DNA to its undamaged state (Kaina et al., 2007).

1.5.2. Single Strand DNA Damage Repair

In the case of DNA damage to nucleotides on a single strand of DNA, the DNA can be repaired using its complementary strand as a template. Three such mechanisms exist to remove damaged nucleotides and replace them with undamaged counterparts; these are base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR).

BER is used to repair base lesions from reactive oxygen species, ionising radiation, some genotoxic chemicals and spontaneous deamination (Almeida and Sobol, 2007; Hitomi et al., 2007). Also, as AP sites and SSBs are formed as intermediates of the

BER pathway, the BER machinery can also repair these lesions. In BER, damaged bases are excised before strand incision, gap filling then ligation.

Damaged bases distort the DNA helix and are recognised by specific DNA glycosylases. There are two types of DNA glycosylases: monofunctional glycosylases, which hydrolyse the glycosidic bond between the base and the deoxyribose sugar before allowing strand incision by AP endonuclease (APE1), and bifunctional glycosylases, which have their own AP lyase activity so can carry out strand incision (Hazra et al., 2007; Hitomi et al., 2007). XRCC1 interacts with glycosylases and acts as a scaffold to bring APE1 to the site of damage (Vidal, 2001). Pol β then performs repair synthesis to insert a base into the gap left by strand incision (Dianova et al., 2004). Finally, a complex formed through the binding of DNA ligase III to XRCC1 seals the DNA.

Lesions that are refactory to short patch repair can undergo long patch BER. In long patch BER strand incision occurs as above then polymerases δ, ε, or β, combined with PCNA, poly(ADP-ribose) polymerase 1 (PARP1) and DNA Ligase I synthesis DNA to fill the gap and, in doing so, displace a DNA flap of up to 13 bases in length (Fortini et al., 1998; Gary, 1999; Prasad, 2001; Prasad et al., 2000; Stucki et al., 1998). FEN1 then breaks down the resulting flap and DNA Ligase I ligates the DNA.

Unlike BER, NER doesn't have a set of specific enzymes that detect different lesions, rather the distortion caused in the backbone of DNA by lesions; consequently, NER recognises a huge range of DNA damage. NER is used to remove bulky DNA adducts such as those caused by BPDE, nitrosamines, crosslinking agents such as cis-platin, and UV lesions including CPDs and (6-4)PPs (Hess et al., 1997; Kuraoka et al., 2000; Pfeifer, 2008; Reardon et al., 1999). During NER, a damaged strand is identified and bound, before incision and removal of a stretch of the damaged DNA strand spanning the lesion and filling in of this stretch using replicative enzymes. NER proteins are named for their associations with hereditary diseases, the xeroderma pigmentosum group A-G proteins (XPA-G) and Cockayne Syndrome proteins A and B (CSA and CSB).

NER begins through the preferential binding of XPC, along with HR23B and centrin2, to distorted DNA (Araki et al., 2001; Sugasawa et al., 1998). The DNA-damage binding

complex, composed of DDB1 and XPE, also binds damaged DNA and this step is more critical in the repair of damage that is not so distorting of the DNA e.g. CPDs (Tang and Chu, 2002). A multiprotein transcription and repair complex, transcription factor IIH (TFIIH) is then recruited to the damaged site at the XPC/HR23B/centrin2 complex (Volker et al., 2001). Subsequent recruitment of XPA and RPA prompts the detachment of the cyclin-dependent kinase-activating kinase (CAK) subcomplex of TFIIH. Detachment of CAK and the subsequent recruitment of XPG cause the unwinding of duplex DNA and release of XPC/HR23B/centrin2 (Coin et al., 2008). An XPF/ERCC1 complex then cuts in the 5' direction, and XPG cuts in a 3' direction, to produce an ssDNA region of 24-32 nucleotides in length (J. C. Huang et al., 1992; Matsunaga et al., 1996; O'Donovan et al., 1994). The resulting gap is filled in by DNA polymerases δ or ε, PCNA and RPA and is ligated by either DNA ligase I or a complex of XRCC1 and Ligase III (Moser et al., 2007; Popanda and Thielmann, 1992; Shivji et al., 1992).

NER can also be coupled with transcription. RNA polymerase II stalls when it reaches damaged DNA and recruits ATP-dependent chromatin remodelling factor CSB. CSB then recruits a number of additional factors including the p300 histone acetyltransferase TFIIH, RPA, a CSA-DDB1 complex and other core NER proteins prior to incision and gap-filling as above (Fousteri et al., 2006; Svejstrup, 2002).

The MMR pathway can confer an additional 1000-fold level of fidelity to replication (Hsieh and Yamane, 2008). MMR begins through the recognition of mismatches; eukaryotic MutS homologues recognise DNA mismatch. MutS binds mismatched DNA with an affinity of between 10 and >1500 times higher than regular DNA (Gradia, 2000; Schofield et al., 2001). MMR is an ATP-dependent pathway and 2 MutS subunits in a heterodimer are required for ATPase activity (R. R. Iyer et al., 2006). This MutS heterodimer is joined by a heterodimer of eukaryotic MutL homologues to initiate MMR (R. R. Iyer et al., 2006). The DNA is then nicked either on either the 5' or 3' side of the mismatch and the DNA is excised in a bidirectional manner by exonuclease I and terminates at around 150 base pairs from the mismatch (Dzantiev et al., 2004; R. R. Iyer et al., 2006). RPA, RFC, PCNA and pol δ are then recruited to allow resynthesis in this region (R. R. Iyer et al., 2006).

1.5.3. DNA Double Strand Break Repair

There are a number of DSB repair pathways including the non-homologous end-joining (NHEJ) and the homologous recombination (HR) pathways. Whilst NHEJ mediates the error-prone direct re-ligation of DNA DSBs and is not restricted to any particular cell cycle phases, the more error-averse HR requires a homologous template from the sister chromatid and is only active during the S and G2 phases, due to availability of the sister chromatids.

NHEJ is initiated through the binding of a Ku heterodimer to the DSB (Mari et al., 2006; Walker et al., 2001). Ku is a ring-like structure that quickly associates to DSBs as a result of its high relative abundance and its high affinity for DNA ends ($K_m = 2.4 \times 10^{-9}$ M) (Blier et al., 1993; Mimori et al., 1986; Walker et al., 2001). Ku binds to the sugar-phosphate backbone and, as such, is sequence independent (Walker et al., 2001). Ku then serves as a scaffold to recruit the other NHEJ proteins, including DNA-PKcs, X-ray cross complementing protein 4 (XRCC4), DNA Ligase IV, XRCC4-like factor (XLF) and Aprataxin-and-PNK-like factor (APLF) to DSBs (Costantini et al., 2007; Grundy et al., 2013; Mari et al., 2006; Uematsu et al., 2007; Yano et al., 2007). XRCC4 interacts with Ku and forms superhelical filaments with XLF that are able to bridge broken DNA strands (Andres et al., 2012; Hammel et al., 2011).

Once brought together, the two ends then require processing to make them ligatable. Aprataxin and polynucleotide kinase/phosphatase (PNKP) are involved in the removal of end-blocking groups (Ahel et al., 2006; Bernstein et al., 2005). APLF, Werner and artemis are all proteins responsible for the resection of DNA ends (Kanno et al., 2007; Kusumoto et al., 2008; Perry et al., 2006; Povirk et al., 2007). Any gaps that need filling are performed by the family X polymerases μ and λ (Nick McElhinny et al., 2005). All of these processing enzymes bind to the Ku/XRCC4 scaffold. The final step in NHEJ is the ligation of the broken ends by DNA ligase IV (Grawunder et al., 1997). Ubiquitination of Ku then mediates its dissociation from the now-continuous DNA strand (Feng and J. Chen, 2012). By virtue of its inherent propensity for generation of mutations, NHEJ is also used in V(D)J recombination, the mechanism used to create a diverse collection of T and B cells (Reviewed in (Malu et al., 2012)). Unlike homolgous recombination, NHEJ is utilised across all three domains of life, including bacteria and archaea, as well as the eukaryotic mechanism described here (Bartlett et al., 2013; Della et al., 2004; G. R. Weller et al., 2002).

HR is used for the restoration of DSBs and removal of interstrand crosslinks, as well as being vital for telomere maintenance, chromosome segregation and maintaining the karyotype in meiosis. Mutations in the HR enzymes cause extreme sensitivity to ionising radiation in *S. cerevisiae* and, in humans, can cause heritable diseases such as Fanconi anaemia, a cancer-prone disease, and familial breast cancer (Game and Mortimer, 1974; Kennedy and D'Andrea, 2006).

Homologous recombination is stimulated by the production of a DSB and is initiated through the binding of the MRN (Mre11, Rad50, Nbs1) complex either side of the break (van den Bosch et al., 2003). The ends are subsequently resected by nucleases, CtIP, and Dna2 in concert with the BLM helicase to give ssDNA tails with 3'-OH ends (Mimitou and Symington, 2009). DNA pairing and strand exchange is then carried out by the Rad51 recombinase; this process is stimulated by Rad52 and RPA (New et al., 1998). Rad51 is an essential gene that exists as a homohexamer in cells; it assembles onto ssDNA to form helical presynaptic filaments, assisted by BRCA2 (X. Yang et al., 2005). The presynaptic filament then binds and stretches a duplex DNA molecule at random points to search for homology (Klapstein et al., 2004). Once homology is found, strand invasion occurs and the presynaptic filament becomes topologically intertwined with the duplex DNA, in the process creating a displacement loop (D Loop).

Following strand invasion, HR can take one of two distinct routes – double strand break repair (DSBR) and synthesis-dependent strand annealing (SDSA) (Helleday et al., 2007). In DSBR, the second DSB end is then captured through annealing to the extended D loop and a double Holliday junction is formed. The resolution of the Holliday junction through the action of nicking endonucleases results in both crossover and non-crossover products (Helleday et al., 2007). In SDSA, however, there is no association with crossover products. The SDSA invading 3' strand is extended by a DNA polymerase and then released from the D-loop. This strand is then able to anneal to the second end of the DSB and can be ligated following processing (Helleday et al., 2007).

1.6 DNA Damage Tolerance

The replication fork proceeds at a very high fidelity and processivity, it needs to if it is to faithfully replicate all 3 billion base pairs of the human genome. However, there is a flip side to this intricate replication machinery in that it is very sensitive to damage on the

template it is copying. If DNA damage persists into S phase, the replication fork will stall at these lesions. DNA damage tolerance mechanisms exist to bypass these lesions during S phase. There are three major types of DNA damage tolerance mechanisms and these include firing of dormant origins downstream of the lesion, recombination-mediated restart and template switching, and Y or B family polymerase-mediated translesion DNA synthesis (TLS) using specialised polymerases (Figure 1.13).

Replication origins initiate a pair of bidirectional forks when they fire. If one of a pair of converging forks stalls at a DNA lesion, the second incoming fork can compensate this. However, if both of a pair of replication forks stall at separate DNA damage sites, the intervening DNA is not replicated. As discussed in Section 1.2.1., not all origins are fired in a round of DNA replication, many lie unused or 'dormant'. Although origins are licensed before S phase, if two adjacent forks have stalled then the cell is able to licence additional forks that were previously dormant (Blow et al., 2011). Whilst the regulation of this mechanism remains unknown one hypothesis is that the ataxia telangiectasia and Rad3-related kinase (ATR), which is activated at stalled forks, can phosphorylate the MCM helicase to promote the initiation of dormant origins (Blow et al., 2011).

Replication can be restarted by a recombination-mediated mechanism, promoted by DNA helicases such as Bloom syndrome helicase or Werner syndrome ATP-dependent helicase (Machwe et al., 2006). This involves a fork regression after the replication machinery has stalled, forming a so-called 'chicken foot' Holliday junction structure, followed by one of three proposed mechanisms (Petermann and Helleday, 2010). Firstly the double-stranded DNA end of the chicken foot Holliday junction can undergo strand invasion back into the template DNA to form a D loop with the assistance of Rad51 (Petermann et al., 2010). This D loop allows re-loading of the replication machinery but a double Holliday junction structure is formed, this structure then requires dissolution, which avoids recombination products being generated (Petermann and Helleday, 2010). If stalling persists for an extended period, the replication fork collapses and fork-associated double-strand breaks can be generated by the crossover junction endonuclease MUS81-EME1 through cleavage of the chicken foot Holliday junction (Hanada et al., 2007; 2006). This can then be restored by homologous recombination and a new replication fork can be begun by dormant origin

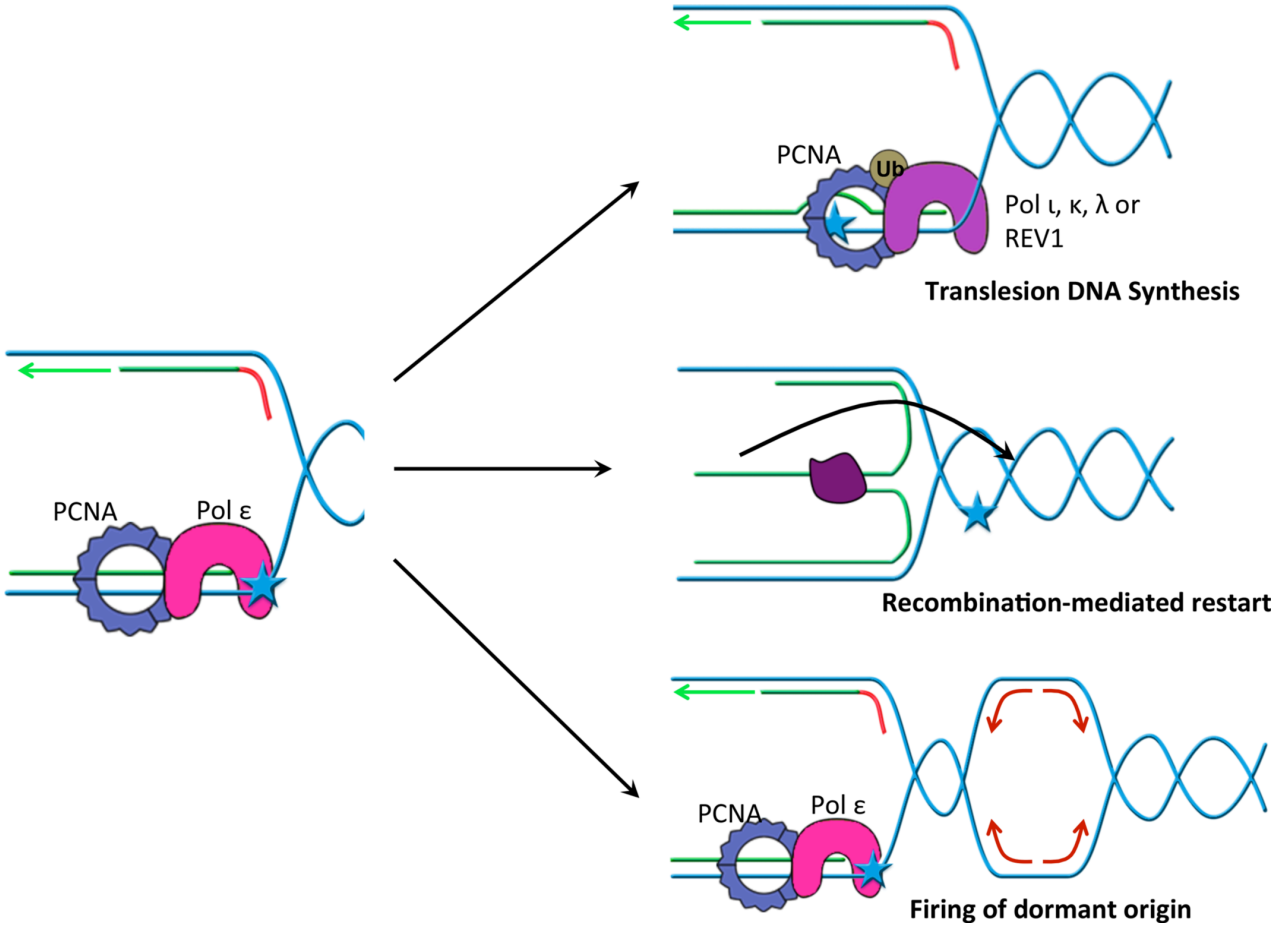


Figure 1.13: DNA damage tolerance mechanisms.
Cells harbour a number of mechanisms for tolerating DNA damage during replication. These include the bypass of lesions by specialised DNA polymerases known as translesion DNA synthesis (TLS); a mechanism of recombination-mediated restart following the production of DNA in a 'chicken foot' structure; and the firing of a downstream dormant origin.

firing in order to replicate the intermediate DNA (Petermann and Helleday, 2010). Finally, it has been proposed that restart could occur through fork remodelling; this would include the reversal or degradation of the chicken foot Holliday junction (Petermann and Helleday, 2010). However, fork restart by remodelling would require dissociation of the replication machinery and the machinery would need to be loaded back onto the DNA and it remains unclear how this would occur.

TLS involves the use of specialised, low-fidelity polymerases that can extend nascent DNA chains over damaged DNA bases. TLS polymerases include pols η, ι, κ and REV1 from the Y family, pol μ from the X family and pol ζ from the B family of polymerases (Hübscher and Maga, 2011). These polymerases maintain little sequence homology with replicative polymerases, however they share a similar overall structure. Like replicative polymerases, these TLS polymerases' catalytic domains resemble a right hand in topology, with fingers, thumb and palm domains (Biertümpfel et al., 2010; Ling et al., 2001). In contrast, unlike replicative polymerases, they have a larger active site capable of accommodating bulky adducts. Some also contain a little finger domain, responsible for making DNA contacts close to the site of the lesion. As a different polymerase is required to replicate through the lesions, there must exist a mechanism for switching the polymerases from the replicative polymerase to the lower-fidelity TLS polymerases. There are two essential mechanisms for the regulation of TLS, one of which occurs during replication and requires the non-catalytic function of REV1 and one which occurs postreplicatively and requires the monoubiquitination of PCNA.

Following UV irradiation, the processivity factor PCNA becomes ubiquitinated and this ubiquitination linkage is formed via Lys-63, rather than the more common Lys-48 ubquitin linkage that directs proteins for degradation (Hoege et al., 2002; Kannouche et al., 2004). Specialised ubiquitin binding domains on TLS polymerases (UBM or UBZ domains) increase their affinity for ubiquitinated PCNA (Lehmann et al., 2007). The current prevalent theory for which polymerase is chosen is that the polymerase is selected by mass action, if ubiquitinated PCNA is located at a lesion, one of the TLS polymerases will bind and attempt to bypass the lesion and subsequently, if unable to do so, another polymerase will attempt to bypass the lesion (Lehmann et al., 2007). As REV1 is able to interact with all of the other TLS polymerases, combined with the fact that removal of its dCMP transferase activity does not abrogate its requirement in yeast for UV-induced mutagenesis, this suggests REV1 may have an additional role in acting

as a scaffold for other TLS polymerases and allowing switching between the two, allowing one polymerase to insert nucleotides opposite a lesion and a second polymerase to extend from such insertion (Friedberg et al., 2005). As TLS often results in mismatches, it may be that the initial TLS polymerase cannot extend from the base it has inserted and it may require pol ζ, which can extend from mismatches to subsequently elongate the chain following insertion opposite the lesion (Acharya et al., 2006). It remains unclear whether PCNA becomes de-ubiquitinated following translesion DNA synthesis and further work is required to establish whether this is the mechanism for regulation of TLS following lesion bypass (Friedberg et al., 2005; Liang et al., 2014).

1.6.1. DNA Damage Tolerance in the Mitochondria

The mitochondrial matrix is a highly oxidative environment. Consequently, the level of oxidative DNA damage, such as 8-oxo-G, as well as AP sites is often high in the mitochondria. Whilst less critical than replication of the nuclear genome due to the mitochondria having multiple DNA copies, it remains important to maintain the mitochondrial genetic code as mutations are associated with aging (Rose et al., 2010). The replicative polymerase of mitochondria, pol γ (see Section 1.2.5) is hindered by most forms of DNA damage. Pol γ deals poorly with 8-oxo-G moieties, and this lesion forms an impediment to DNA replication by pol γ (Garcia-Gomez et al., 2013; Graziewicz et al., 2007). In the case of AP sites, pol γ does show significant inhibition, but is able to incorporate an adenine opposite an abasic site (Garcia-Gomez et al., 2013; Pinz et al., 1995). Pol γ is inhibited by UV lesions such as CPDs; however exo-deficient pol γ is able to bypass these lesions with low efficiency, choosing to insert guanine opposite the 3'-thymine of the CPD (Kasiviswanathan et al., 2012). In 2006, Zhang and colleagues indicated that *REV* mutant strains of *S. cerevisiae* exhibited a reduction in UV-induced frameshifts in the mitochondria (H. Zhang et al., 2006). Additionally, this group also showed that the N-termini of REV1 and Pol ζ were capable of directing GFP to the mitochondria (H. Zhang et al., 2006). Taken together, along with more recent evidence that REV-deficient strains of *S. cerevisiae* result in a 20-fold increase in UV-induced mitochondrial point mutations (Kalifa and Sia, 2007), suggests that there is also an alternative pathway of DNA damage tolerance in mitochondria to bypass UV lesions involving REV1 and Pol ζ.

1.7 DNA Primases

As previously remarked, DNA polymerases, the proteins that carry out the bulk of DNA replication, are unable to initiate the *de novo* polymer synthesis reaction. As early studies could find no evidence that DNA strands were initiated by DNA polymerases, Kornberg and colleagues postulated that DNA replication initiation required RNA transcription (Brutlag et al., 1971). This two-step mechanism of replication was then confirmed the next year: the first step requires a ssDNA template, ribonucleotides (rNTPs), and enzymes in a reaction that is rifampicin-sensitive; this is followed by a second step that requires dNTPs and enzymes in a rifampicin-resistant reaction (Wickner et al., 1972). This first step became known as 'priming'. Soon after, the first enzyme capable of priming, later to come known by the term 'primase', was discovered in *E. coli* as the product of the dnaG gene (Bouché et al., 1975).

For decades, the perception of primases existed as specialised DNA-dependent RNA polymerases that were involved in replication initiation. However, more recently a body of evidence has established that primases actually represent a much broader and functionally diverse family of polymerases with a variety of nucleotidyl transferase activities and this will be discussed further in this section.

1.7.1. Evolution of DNA Primases

A number of mechanisms have evolved for the generation of primers (see Table 1.2), but the overwhelming majority of systems utilise the specialised DNA primases to produce short *de novo* nucleotide strands opposite a templating strand that can subsequently be extended by DNA polymerases (Frick and Richardson, 2001). The characteristics of archaeal genomes are similar to that of bacteria: they arrange their genes into operons, their DNA is not contained in a nucleus, chromosomes are circular and nearly all their mRNAs lack introns. However, the DNA replication machinery of archaea distinctly resembles eukaryotes more than bacteria (Edgell and Doolittle, 1997) and this extends to DNA primases. Primases are differentiated on the basis of their structural features, whether they contain an archaeal/eukaryotic fold (L. M. Iyer et al., 2005) or a toprim architecture found in bacterial and phage primases, such as DnaG (Aravind et al., 1998). The superfamily of proteins that contain an archaeal/eukaryotic fold are known as archaeo-eukaryotic primases, or AEPs. This pattern of little or no homology between the achaeal/eukaryotic and bacterial primases extends to other replicative proteins of these domains, including polymerases and

Mechanism of Priming	Utilised In
An endogenous host tRNA is used as a primer for converting the RNA genome into DNA (Mak and Kleiman, 1997).	Retroviruses
Instead of using a specialised primase, the mitochondrial RNA Polymerase produces short RNA transcripts that are then elongated by pol γ (Pham et al., 2006).	Mitochondria
A serine from the terminal protein forms a phosphodiester bond with dCMP at the 5' side of the dCMP, which can be elongated at its 3' side by DNA polymerase (Pronk and van der Vliet, 1993).	Adenovirus
Rolling circle replication (RCR) is utilised by a number of DNA viruses. RCR endonuclease nicks one of the DNA strands. The free 3'-OH is elongated by a DNA polymerase whilst the 5'end of the nicked strand is transferred onto a tyrosine residue of the nuclease (Noirot-Gros and Ehrlich, 1996).	Several DNA virus families (including parvoviruses, geminiviruses and circoviruses) (Ilyina and Koonin, 1992)
A specialised enzyme, known as a primase, is utilised to produce de novo nucleotide strands opposite a template strand at an initiation site (Frick and Richardson, 2001).	-All other known systems including all cellular systems

Table 1.2: Mechanisms of generating primers.
As polymerases are unable to synthesis *de novo* nucleotide polymer strands, different mechanisms have evolved in order to provide a free 3'-hydroxyl for polymerases to extend from using deoxyribonucleotides to initiate DNA replication.

helicases (Edgell and Doolittle, 1997). There are two major theories as to why there are two distinct groups of replication proteins (Figures 1.14-1.15).

The first model suggests that, as a result of divergence of sequence but convergence of function, DNA replication, or at least the enzymes involved in nascent strand production, may have evolved twice independently (Leipe et al., 1999) (Figure 1.14). Despite the replicative machinery proteins of bacteria and archaea/eukaryotes being highly divergent, the core components of the transcription and translation machinery including DNA ligases, PCNA, RNAse HII, clamp-loaders, the DNA-dependent RNA polymerase etc are conserved across domains (Leipe et al., 1999; Sweetser et al., 1987). Leipe, Aravind and Koonin hypothesise that the two replicative systems evolved independently from a common ancestor that carried out reverse transcription to replicate an RNA/DNA genome (Leipe et al., 1999). Reverse transcription is the process by which an RNA genome would be reverse-transcribed into an RNA/DNA heteroduplex, the RNA strand would then be degraded by a protein with RNase H nuclease activity and the resulting single-stranded DNA would act as a template for the synthesis of the double stranded DNA; RNA would then subsequently be transcribed by a DNA-dependent RNA polymerase for protein production. Following the evolution of DNA replication-competent cells, selective pressure would result in the elimination of the reverse transcription pathway. This is supported by the evidence that a number of primases and polymerases inherently, or can be modified to, exhibit reverse transcriptase activity (Gill et al., 2014; Jozwiakowski and Connolly, 2011; Myers and Gelfand, 1991; Ong et al., 2006). This model suggests that the transcription in this last universal common ancestor would occur more directly using the RNA/DNA genome and, as such, this would account for the conservation of core components of the transcription and translation machinery between archaea/eukaryotes and bacteria. There remains a number of reverse transcribing genetic elements including, for example, reverse transcriptases but they are not necessarily indispensible for viability (Greider and Blackburn, 1989).

A second model suggests that the last universal common ancestor employed a dual-primase system (Hu et al., 2012)(Figure 1.15), consisting of both an AEP and a toprim primase. These two groups of primases have both been described in both bacteria (Koonin et al., 2000) and in archaea (Zuo et al., 2010). In addition to a dual-primase role in replication, the AEP and toprim primases could have served roles in NHEJ

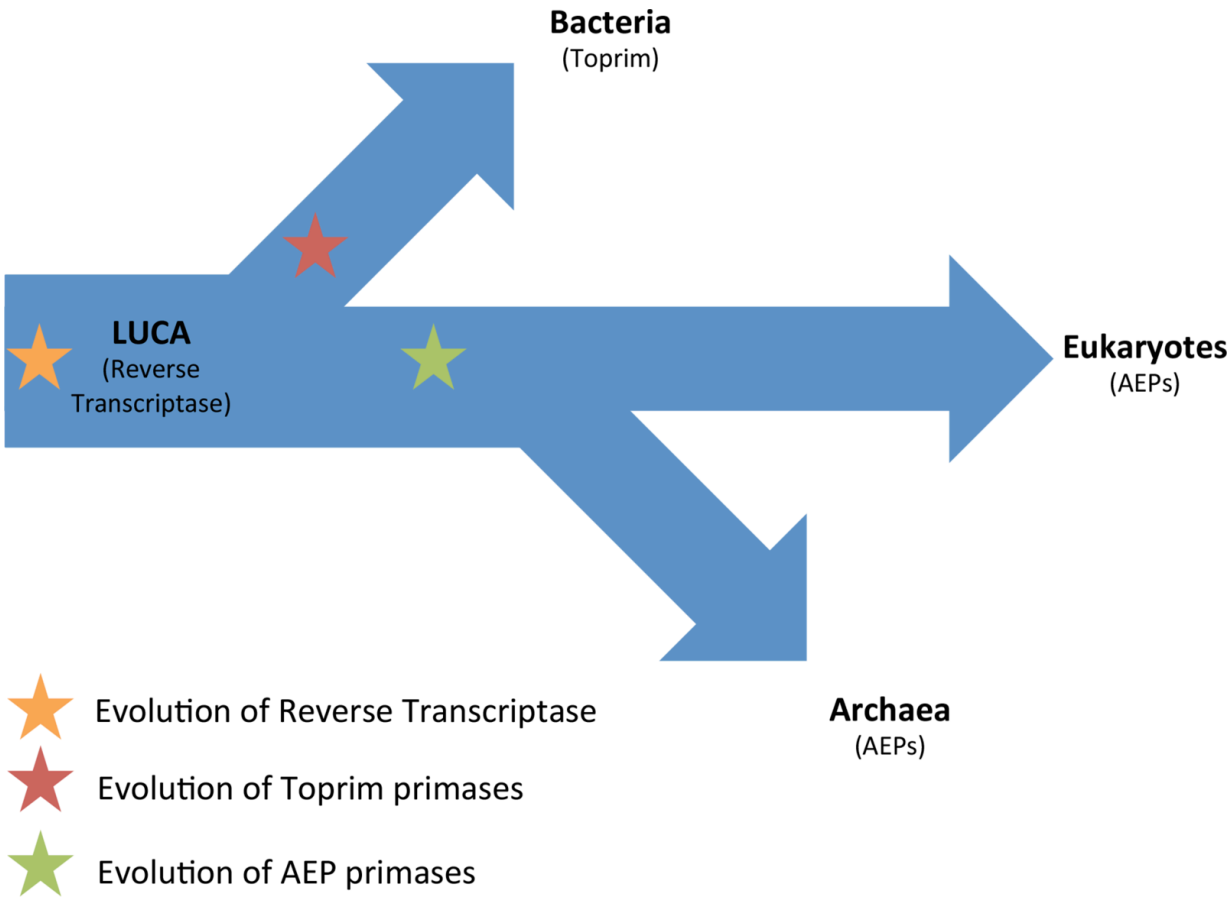


Figure 1.14. Evolution of primases (1).
The first model of primase evolution suggests that primases evolved twice independently from a last universal common ancestor (LUCA). Bacterial ancestors evolved toprim primases and archaeal and eukaryotic ancestors evolved AEP primases. Subsequent horizontal gene transfer occurred between the two lineages to account for AEP primase's role in NHEJ in bacteria and toprim-type primase's role in archaeal RNA degradation.

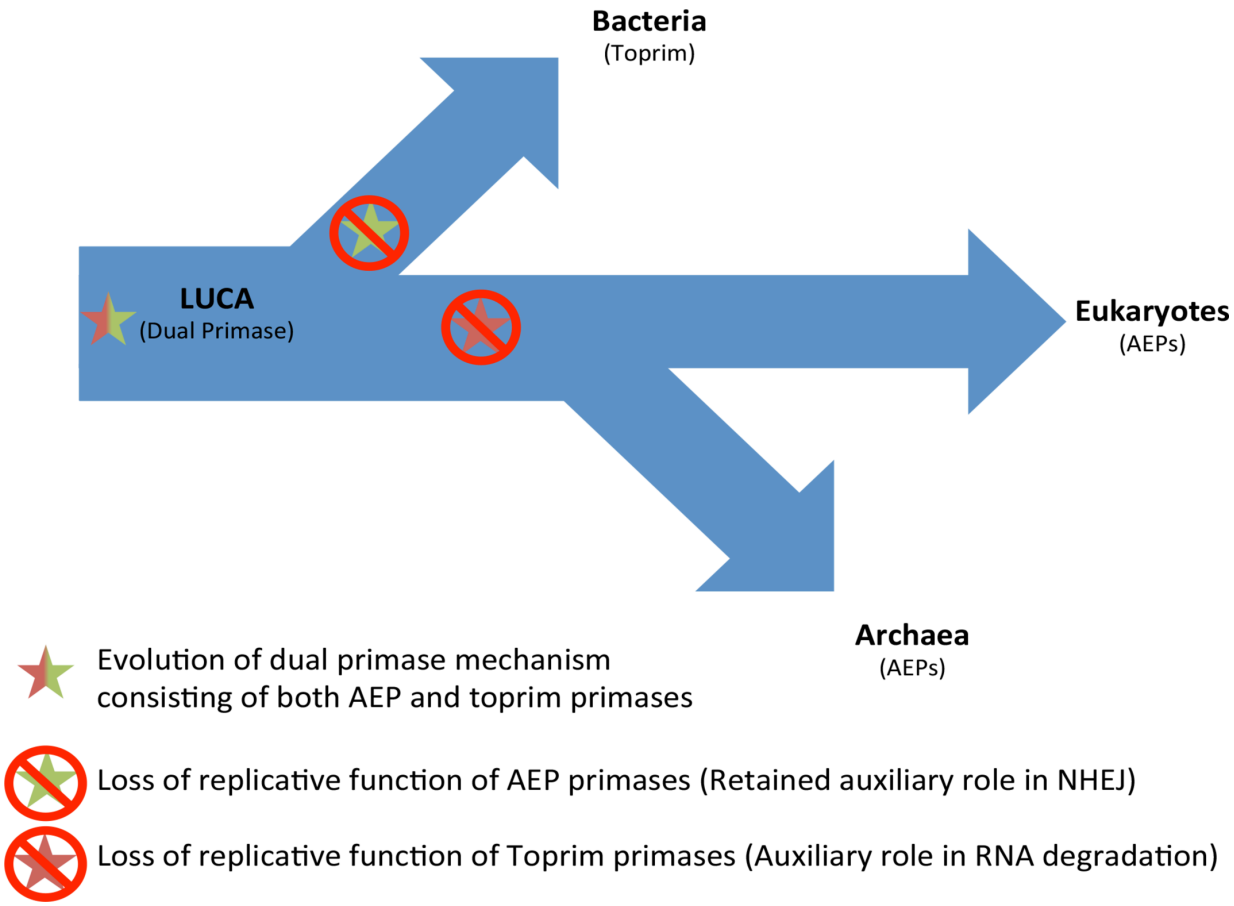


Figure 1.15. Evolution of primases (2).
The second model of primase evolution suggests that the last universal common ancestor (LUCA) had a dual primase replication mechanism, consisting of both AEP and toprim primases. During the evolution of bacteria, they lost the replicative function of the AEP primases but retained them for the auxiliary function of NHEJ. During the evolution of the archaeal/eukaryotic lineage, the replicative function of toprim primases were lost, but their auxiliary role in archaeal RNA degradation was retained.

processes and RNA degradation respectively in the last universal common ancestor. In bacteria selective pressure resulted in the loss of AEPs as replicative primases but retained for their role in NHEJ (Della et al., 2004; G. R. Weller and A. J. Doherty, 2001) and in archaea toprim primases were lost as a replicative primase, but still retained for their non-essential role in RNA degradation (Le Breton et al., 2007; Walter et al., 2006). This model predicts that, in eukarya, toprim primases are lost and other proteins have evolved to take its place (Hu et al., 2012).

An alternative scenario to these two models could be that the DNA replication machinery of a common ancestor underwent dramatic evolutionary events to cause the dissimilarity in core replicative proteins under unknown, but serious and distinct, evolutionary pressures (Figure 1.16).

1.7.2 The Archaeo-Eukaryotic Primase (AEP) Superfamily of Proteins

Archaeo-Eukaryotic Primases (AEPs) share a catalytic core of an $(\alpha\beta)_2$ unit that has no equivalent structure to any other proteins of the PDB database. This catalytic core harbours three conserved motifs (motifs I, II and III), namely an hhhDhD/E motif (where 'h' is a hydrophobic residue), an sxH motif (where 's' is a small residue and 'x' is any residue) and an hD/E motif (Figure 1.17) (L. M. Iyer et al., 2005). The first and third of these motifs are involved in metal coordination for catalysis; despite divergent sequences in the scaffolds of polymerases and primases, they share a common two metal mode of catalysis (see Section 1.3.2) and the metal ion coordinating core of pol β is superimposable to that of the replicative primase (Augustin et al., 2001; Lao-Sirieix et al., 2005; Lipps et al., 2004). The sxH motif is required for nucleotide binding. Multiple mutagenesis studies have shown these motifs to be essential for catalysis (Augustin et al., 2001; Bianchi et al., 2013; Copeland and Tan, 1995; Della et al., 2004; Galal et al., 2012b; Klinedinst and Challberg, 1994; Lao-Sirieix and S. D. Bell, 2004; Lipps et al., 2003; Mikhailov and Rohrmann, 2002). Almost all primases also have zinc-binding motifs that are required for catalysis and exceptions to this will be discussed later in this section. In addition to the AEP polymerase domain, some AEPs have a number of associated domains, including zinc-binding domains and helicase domains, these are illustrated in Figure 1.18.

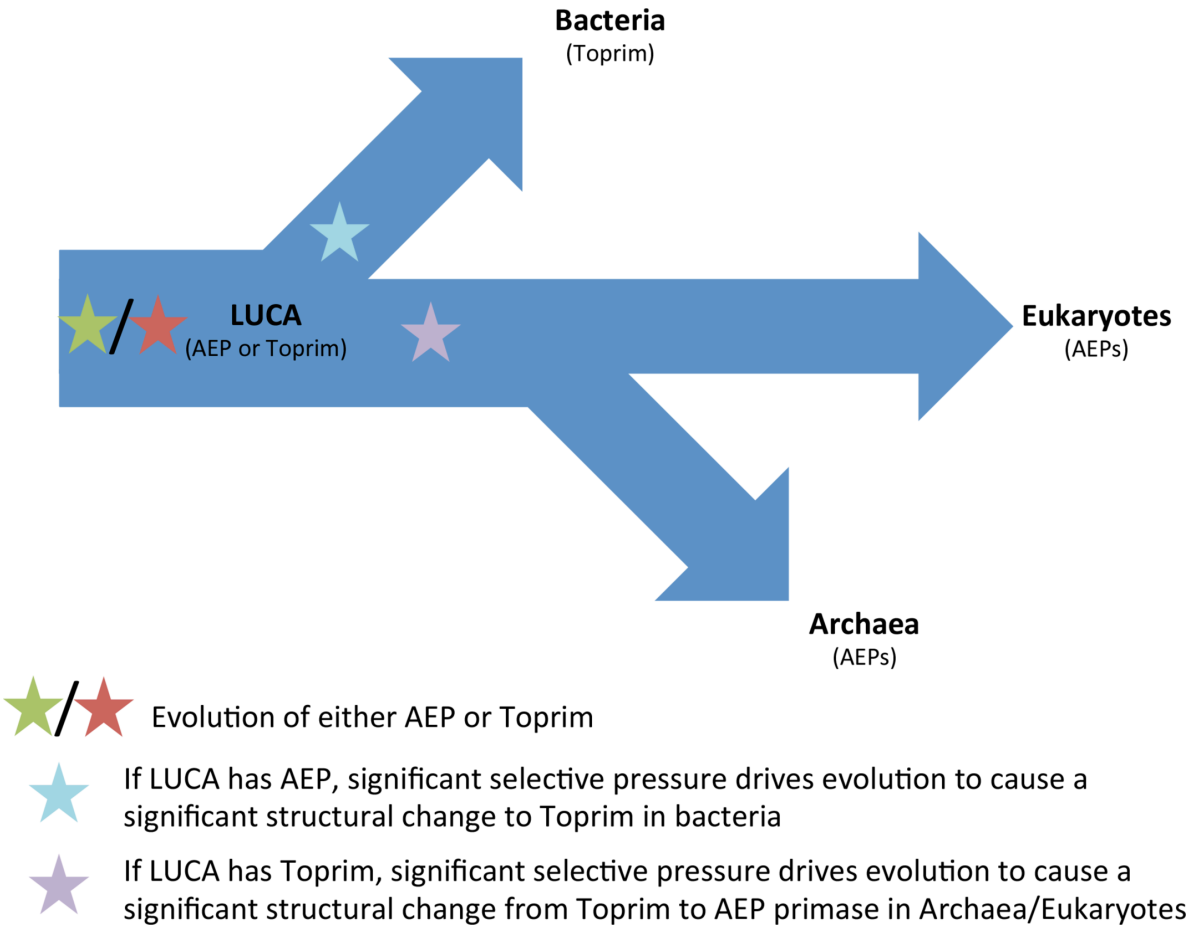


Figure 1.16. Evolution of primases (3).
The third model of primase evolution suggests that the last universal common ancestor (LUCA) had either an AEP or toprim-like primase. Significant, unknown evolutionary pressures could then have driven whichever type of primase was present to the other type.

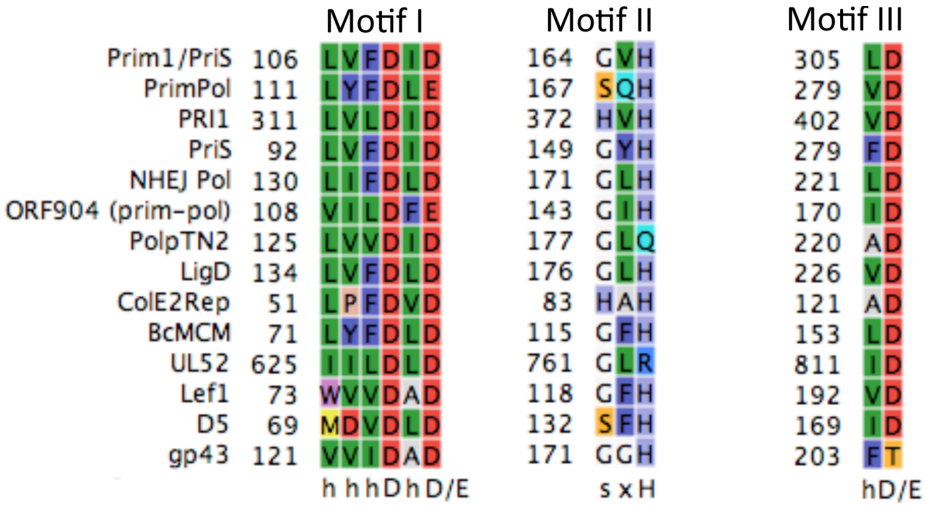


Figure 1.17. Conserved motifs of the AEP superfamily.
The AEP superfamily is formed of three conserved motifs. The first motif is a hhhDhD/E, where ‘h’ is a hydrophobic residue. The second motif is an sxH motif, where ‘s’ is a small residue and ‘h’ is any residue. The third motif is an hD/E motif. The first and third motifs are responsible for metal ion binding in catalysis and the second motif is responsible for nucleotide binding.

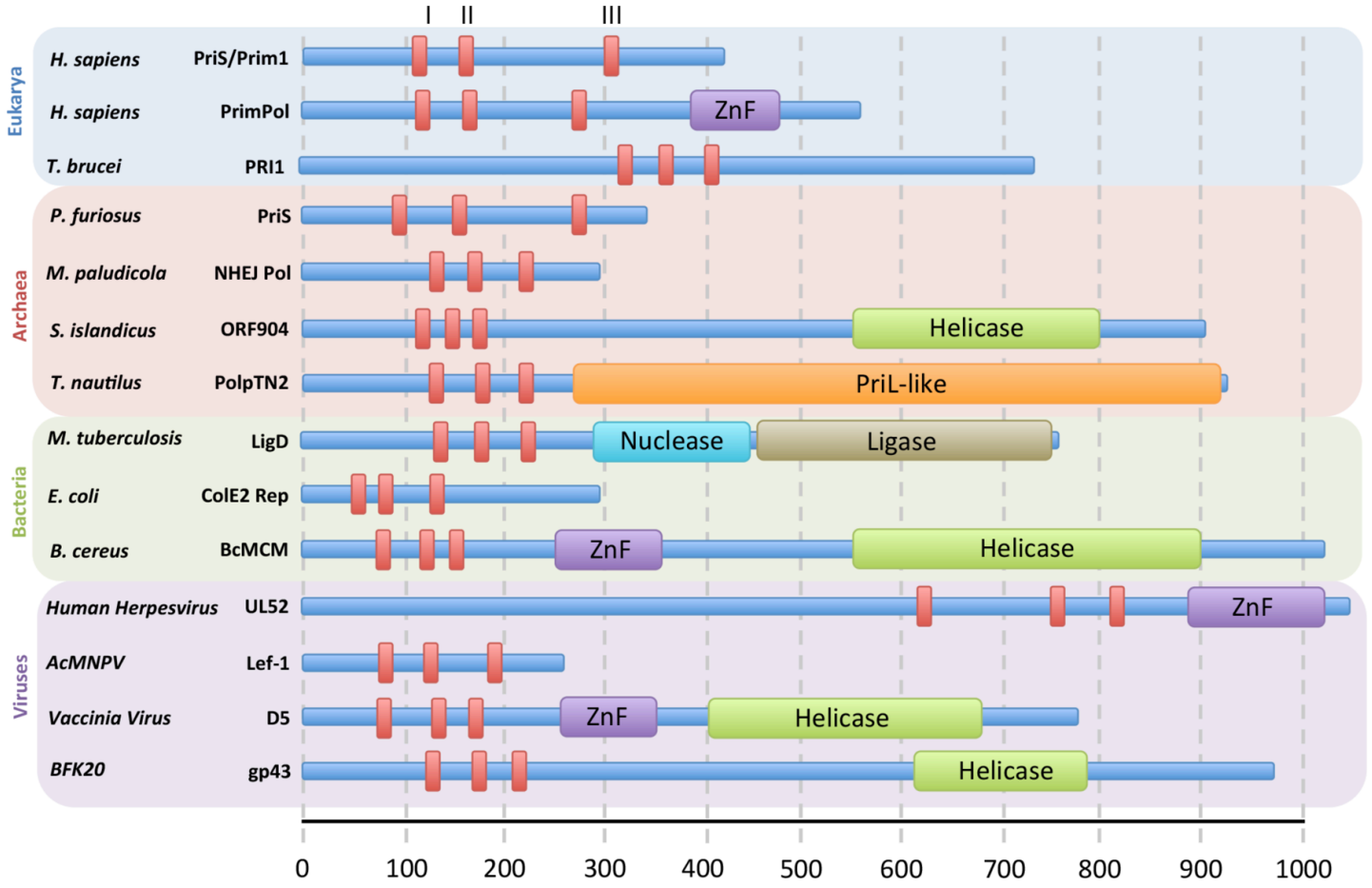


Figure 1.18. Domain organisation of different families of the AEP superfamily.
The AEP superfamily is formed of a number of divergent families with different domain organisations. The three motifs of the AEP polymerase domain are depicted (red) along with any accessory domains of examples of each of these biochemically-characterised families. The blue, red, green and purple backgrounds correspond to the domain of life in which the primase family is found.

There is currently no structural information for the initiation step of any AEPs therefore the exact mechanism of primase activity remains unclear. The current proposition is that a quarternary complex is formed of primase, DNA template and two nucleoside triphosphates; the primase in this case has two incoming nucleotide binding sites (Frick and Richardson, 2001). However, the recent structure of the RNA-dependent RNA polymerase from the Hepatitis C virus may help shed some light on the mechanism proteins use to form dinucleotides and extend from this structure (Appleby et al., 2015). These polymerases are the viral replicative polymerases and, unlike other replicative polymerases, are capable of both *de novo* polymer synthesis and primer extension. A β-loop insertion in the thumb domain of the polymerase acts as a priming platform to buttress the initial priming nucleotide and allow the formation of a dinucleotide. Following this initial step, the β-loop recedes from the catalytic core of the polymerase, supporting and guiding the nascent RNA chain along through an exit channel.

As previously mentioned, there is a growing body of evidence that primases, particularly those of the AEP superfamily, have a range of nucleotidyl transferase activities, these are described in the upcoming sections and are summarised in Figure 1.19 and Table 1.3.

1.7.2.1. Eukaryotic AEP Primases

Primase activity in eukaryotes was first associated with the polymerase α complex as early as 1982 (Conaway and Lehman, 1982b). Pol α consists of a 180 kDa subunit, responsible for the DNA polymerase activity of the complex, a 70 kDa subunit, with no catalytic functions, and two subunits of 49 kDa (Prim1/PriS) and 58 kDa (Prim2/PriL) (T. S. F. Wang, 1991). PriS alone has primer extension activities using ribonucleotides but requires PriL for efficient priming activity (Copeland and T. S. Wang, 1993). The average length of the primers produced by pol α range between 8 and 15 nucleotides (Brooks and Dumas, 1989; Conaway and Lehman, 1982a; T. S. Wang et al., 1984). The primase activity of pol α is of moderate fidelity but this can be reduced through the addition of manganese (Kirk and Kuchta, 1999a). PriL is predicted to be important for the dinucleotide synthesis and translocation activities of the primase as well as regulation of the length of the elongation products of the primase complex (termed 'counting') of pol α (Arezi et al., 1999; Zerbe and Kuchta, 2002). This PriL subunit, in higher eukaryotes, also contains an iron-sulphur cluster that is indespensible for primase activity in its C-terminal region (Klinge et al., 2007; Vaithiyalingam et al., 2010;

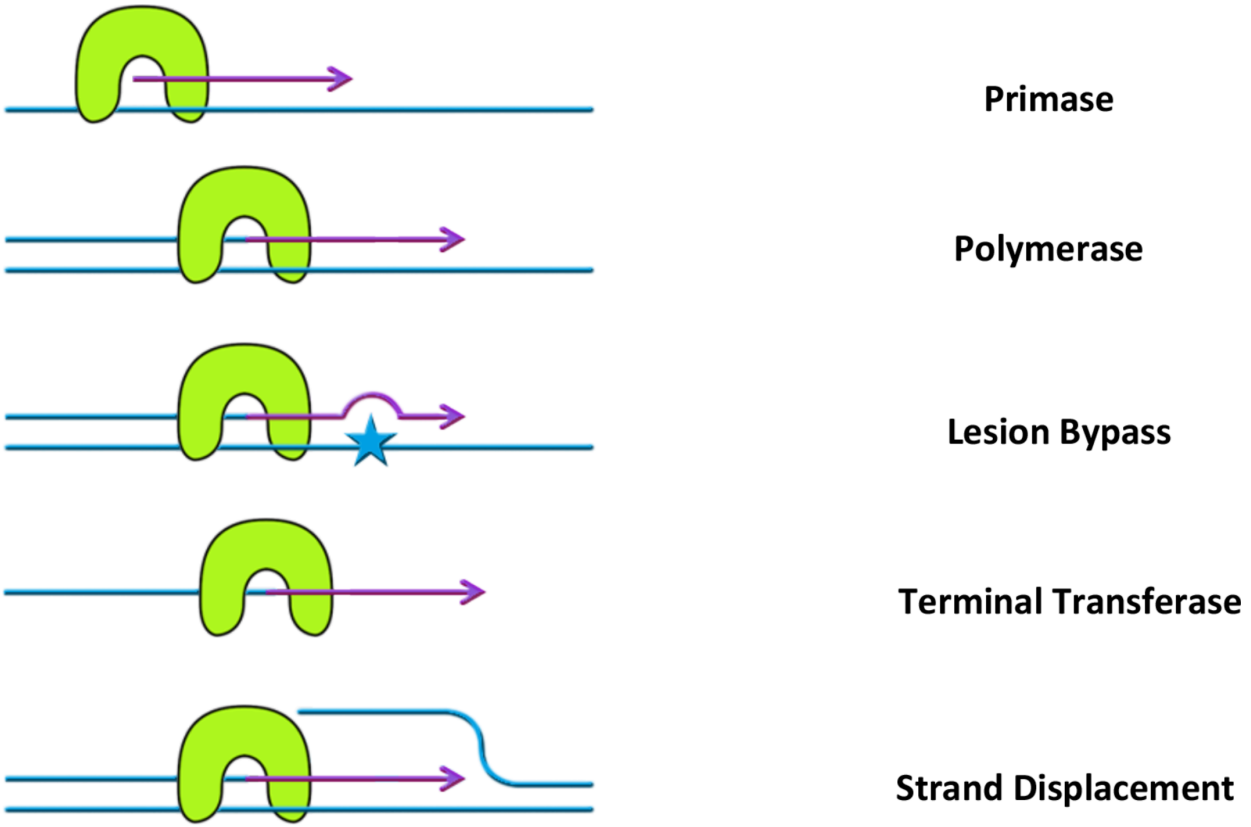


Figure 1.19. Nucleotidyl transferase activities reported in AEP primase-polymerases.
AEP primases have more activities than priming DNA for replication. The reported additional nucleotidyl transferase activities for different AEPs are depicted including polymerase activity (either DNA-dependent DNA polymerase or DNA-dependent RNA polymerase), lesion bypass, terminal transferase and strand displacement. See Table 1.3 for examples of which primases have been shown to have the activities described above.

Domain	Protein Family	Primase	Polymerase	Lesion Bypass	Terminal Transferase	Strand Displacement	References
Eukarya	PriS/Prim1	✔	✔				1-3
Eukarya	PrimPol	✔	✔	✔	✔		4-6
Eukarya	PRI1/2	✔					7-8
Archaea	PriS	✔	✔	✔	✔	✔	9-14
Archaea	NHEJ Pol		✔		✔	✔	15
Archaea	prim-pol (ORF904/Rep)	✔	✔		✔		16-17
Archaea	PolpTN2	✔	✔				18
Bacteria	LigD	✔	✔	✔	✔	✔	19-22
Bacteria	ColE2 Rep	✔					23-24
Bacteria	BcMCM	✔	✔				25
Viral	UL52	✔					26-27
Viral	Lef-1	✔	✔				28
Viral	D5	✔	✔				29
Viral (phage)	gp43	✔	✔				30

References: 1. Brooks and Dumas, 1989; 2. Conaway and Lehman, 1982; 3. Copeland and Wang, 1993; 4. Bianchi et al., 2013; 5. Garcia-Gomez et al., 2013; 6. Keen et al., 2014; 7. Hines and Ray, 2010; 8. Hines and Ray, 2011; 9. Desogus et al., 1999; 10. Liu et al., 2001; 11. De Falco et al., 2004; 12. Le Breton et al., 2007; 13. Hu et al., 2012; 14. Jozwiakowski et al., 2014; 15. Bartlett et al. 2013; 16. Lipps et al., 2003; 17. Prato et al., 2008; 18. Gill et al., 2014; 19. Della et al., 2004; 20. Pitcher et al., 2005; 21. Gong et al., 2005; 22. Pitcher et al., 2007; 23. Takechi and Itoh, 1995; 24. Takechi et al., 1995; 25. Sanchez-Berrondo et al., 2012; 26. Crute and Lehmann,1991; 27. Klinedinst and Challberg, 1994; 28.Mikhailov and Rohrmann, 2002; 29. De Silva et al., 2009; 30. Halgasova et al., 2012.

Table 1.3. Nucleotidyl transferase activities of biochemically characterised AEPs.
The reported nucleotidyl transferase activities from Figure 1.18 for AEPs that have been biochemically characterised. It should be noted that the primases listed here have not necessarily been tested for all of these activities.

Weiner et al., 2007); this motif remains absent from the archaeal proteins that have thus far been investigated. The crystal structure of this iron-sulphur cluster reveals that this motif is integral to the structure of PriL and contributes to the C-terminal region's binding to primer-template DNA (Vaithiyalingam et al., 2010). However, given the complexity of an iron-sulphur cluster it is unlikely that the role of this motif has evolved as a purely structural one and it may have additional uncharacterised biochemical activities; it has recently been reported to have an undefined role in the loading of the polymerase complex at replication origins, and that pol α loading is affected by the redox state of the cell (L. Liu and M. Huang, 2015). It should also be noted that iron-sulphur clusters, whilst their biochemical activities remain largely unknown, are present in other DNA processing proteins of DNA repair pathways such as DNA glycosylases of BER and DNA helicases of NER and have been suggested to harbour the ability to discern damaged DNA (Fan et al., 2008; Lukianova and David, 2005).

Recently, a second family of primases was discovered in higher eukaryotes, Primase-Polymerase (PrimPol). This primase is introduced extensively in Section 1.8 and is the main topic of this thesis.

1.7.2.2. Archaeal AEPs

Unlike the eukaryotic replicative primase, whose activities are limited to priming and primer extension using ribonucleotides, AEPs of the archaeal domain have an array of activities (Bocquier et al., 2001). The first archaeal primase to be described biochemically was the replicative PriS-like primase from the euryarchaeon *Methanococcus jannaschii* (Desogus et al., 1999). This primase showed the ability to prime opposite homopyrimidine ssDNA templates poly(dT) and poly(dC) (Desogus et al., 1999).

This initial characterisation only looked at the ribonucleotide priming activity opposite DNA templates. Subsequent investigations into the deoxyribonucleotide priming activity of a range of archaeal primases has revealed that these proteins are actually capable of producing DNA primers with varying efficiency. The primase from *Thermococcus kodakaraensis* preferentially primes using rNTPs, with a binding constant about 2-4 fold lower than dNTPs (Galal et al., 2012b). However, this is the only replicative archaeal primase reported to preferentially incorporate rNTPs; *Pyrococcus abyssi* had binding constants for rNTPs and dNTPs in the same range (Le Breton et al., 2007), but

primases from other archaeal species *Pyrococcus horikoshii, Solfolobus solfataricus* and *Pyrococcus furiosus* all have a preference for priming using dNTPs over rNTPs (Lao-Sirieix and S. D. Bell, 2004; L. Liu et al., 2001; Matsui et al., 2003).

Unlike the replicative primase of pol α, archaeal primases are capable of efficiently extending primers using both rNTPs and dNTPs, with a marked preference for dNTPs. These DNA primer extension products can range in size from less than 500 bases in length to over 7 kilobases (Bocquier et al., 2001; Galal et al., 2012b; Lao-Sirieix and S. D. Bell, 2004; Le Breton et al., 2007; L. Liu et al., 2001). This has lead to the suggestion that not only do these replicative primases act as the PriS/L primase of pol α but also the 180 kDa polymerase subunit of pol α (L. Liu et al., 2001); instead of undergoing a mechanism in which the initial primer is handed over to a separate subunit in which pol α polymerase activity can then extend from this short primer before a replicative polymerase takes over, the primers produced by these primases are sufficient for archaeal replicative polymerases to extend. This DNA polymerase activity of many AEPs has lead to the recommendation that AEPs should be identified as primase-polymerases.

The primases from *M. jannaschii* and *S. solfataricus* have both been shown to be capable of terminal nucleotidyl transferase activity at the 3' end of synthetic oligonucleotides in the presence of manganese (De Falco et al., 2004; Lao-Sirieix and S. D. Bell, 2004). This terminal transferase activity allows archaeal primases to undergo template-dependent polymerisation across discontinuous templates (Hu et al., 2012). Consequently, it is proposed that these primases could serve a role in double-strand break repair in archaea. The evidence that Pyrococcus abyssi has strand-displacement and gap-filling activities reinforces this notion that archaeal replicative primases could have a function in archaeal DNA damage repair (Le Breton et al., 2007). Given that archaea lack family X polymerases, that pols β and λ also possess strand-displacement and gap-filling activities, and that the family X polymerases are related to primase (see Section 1.7.2.1) it is not implausible that archaeal replicative primases could have a role in DNA damage repair.

Whereas some archaeal species possess the Y family polymerase paralogue Dpo4, many do not have any Y family polymerases (Kelman and White, 2005). Two such species lacking Dpo4 are *Archaeoglobus fulgidus* and *P. furiosus*, it has recently been

shown through work from our lab that the replicative primases from these species are DNA damage tolerant (Jozwiakowski et al., 2015). A well-documented feature of archaeal replicative polymerases is their ability to discern, and stall at, uracil base adducts (Fogg et al., 2002) caused by the deamination of cytosine (see Section 1.4.5). This cytosine deamination is particularly relevant in the extreme conditions faced by some of the thermophilic as it is a process that is considerably sped up by heat (Lindahl and Nyberg, 1974). The primases of *A. fulgidus* and *P. furiosus* are both capable of replication through uracil bases and can rescue a replicative polymerase bound to PCNA stalled at a uracil base (Jozwiakowski et al., 2015). Additionally, these primases were also capable of bypass of an 8-oxo-G oxidative lesion (Jozwiakowski et al., 2015). Somewhat surprisingly, both of these archaeal primases were also capable of the error-free bypass of CPDs, inserting two adenines opposite the two thymines of this distortive UV lesion.

Linked with this inherent tolerance of archaeal DNA primases is a curious finding regarding the *T. kodakaraensis* replicative primase. Independent of a DNA template, the primase from *T. kodakaraensis* was able to produce discrete bands that migrated to corresponding chain lengths of around 16 and 25 nucleotides (Galal et al., 2012b). NMR and LC-MS/MS analysis revealed these products to be dNMP-Tris and dNMP-glycerol adducts (Galal et al., 2012a). Frick and Richardson propose a model for synthesis of primers that involves 2 binding sites for NTPs, one initiation site that accommodates the 5' nucleotide and one elongation site that accommodates the 3' nucleotide triphosphate (Frick and Richardson, 2001). In this case it is predicted that tris or glycerol is in the 5' nucleotide site, with the incoming dNTP in the 3' site (Galal et al., 2012a). Although not identified as the same products, similar-sized products were evident in activity tests from another AEP, PolpTN2 (see Section 1.7.2.3) (Gill et al., 2014). These dNMP-Tris and dNMP-glycerol products represent a highly tolerant active site of archaeal AEPs.

1.7.2.3. NHEJ AEP polymerases

Bacterial homologues of PriS were identified through a thorough bioinformatics search (Koonin et al., 2000; G. R. Weller et al., 2002; G. R. Weller and A. J. Doherty, 2001). A subsequent bioinformatics study, exploring homology of ku in bacteria identified that bacterial homologues of ku were either operonic with, or located close to, these bacterial homologues of PriS, suggesting a novel family of end-joining proteins (Aravind

and Koonin, 2001). A subsequent biochemical analysis of these proteins, designated LigD, revealed that they are indeed involved in NHEJ in bacteria and consist of a ligase domain, nuclease domain and an AEP primase-polymerase domain (PolDom/LigD Pol) (Della et al., 2004; G. R. Weller et al., 2002; G. R. Weller and A. J. Doherty, 2001). This represents the first evidence that AEPs may have additional roles to replication of DNA and expanded on the capabilities of these proteins.

PolDom has an array of nucleotidyl transferase activities and can perform extension of primers in a template-dependent manner using both DNA and RNA, as well as having gap-filling, strand displacement, TLS, and terminal transferase activities (Della et al., 2004; Gong et al., 2005; Pitcher et al., 2007; 2005; H. Zhu et al., 2006). However, unlike replicative AEPs, PolDom is incapable of producing a *de novo* primer from single nucleotides opposite a DNA template (Brissett et al., 2011). The crystal structure of the DNA synaptic complex of *Mycobacterium tuberculosis* PolDom revealed that PolDom mediates the synapsis of DNA ends that are not necessarily compatible through a variety of interactions, including microhomology base pairing, base mismatches and flipped out bases to form a stable end-joining complex (Brissett et al., 2007).

Recently, an NHEJ repair complex from the archaeon *Methanocella paludicola* was reconstituted *in vitro* (Bartlett et al., 2013). The proteins of this repair complex has activities consistent with their prokaryotic equivalents (Bartlett et al., 2013) and this complex has been hypothesised in a range of euryachaeota, but appears absent in the other main group of archaea, crenarchaeota (Nair et al., 2010). The polymerase (Pol) subunit of this complex is an AEP polymerase capable of gap-filling and strand-displacement with a marked preference for rNTPs over dNTPs. The structure of this polymerase is similar to the prokaryotic NHEJ AEPs (Bartlett et al., unpublished).

Although eukaryotes do not appear to contain specialised AEP polymerases for NHEJ, as is the case in archaea or bacteria, they do share a sequence similarity with the family X polymerases, particularly pol β, both in the catalytic and regulatory regions of these proteins (Kirk and Kuchta, 1999b)(Figure 1.20). Also, despite having a unique structural arrangement, the family X polymerases are superimposable at their catalytic sites with AEPs (Kirk and Kuchta, 1999b; Kuchta and Stengel, 2010). These polymerases exhibit gap-filling, strand-displacement and terminal transferase activity and are involved in repair processes such as base excision repair and indeed NHEJ in

Figure 1.20. Alignment of family X polymerases with DNA primase.
The catalytic regions of human polymerases β, λ and μ are aligned with the small subunit of the pol α primase (Prim1/PriS). The residues highlighted in yellow are the invariant catalytic residues. Those in light blue are conserved charged residues and the residues in green represent conserved hydrophobic regions.

eukaryotes (Yamtich and Sweasy, 2010). More recently, the family A polymerase pol θ has been implicated in a novel DNA repair mechanism, microhomology-mediated end joining (MMEJ). Pol θ is shown to exhibit gap-filling and strand-displacement activities from DNA that has formed synapses at 3' overhangs following resection (Kent et al., 2015).

1.7.2.3 Extrachromosomal DNA Replicative AEPs

In addition to the replicative AEP primase-polymerases, some archaea possess additional AEPs encoded in extrachromosomal plasmids. The first of these to be identified was ORF904 from the pRN1 plasmid of *Sulfolobus islandicus*. This primase is predicted to function in the replication of the pRN1 plasmid. ORF904 represents a novel family of primases present sporadically in crenarchaeal plasmids and Gram-positive bacterial plasmids, called the prim-pol family (L. M. Iyer et al., 2005). ORF904 is a multifunctional enzyme composed of a helicase domain and an AEP primase-polymerase domain, and harbouring ATPase, primase and polymerase activities (Lipps et al., 2003). ORF904 polymerase activity is dNTP-dependent and priming preferentially utilises dNTPs over rNTPs (Lipps et al., 2003). Unlike other AEPs, ORF904 shows sequence specificity to a 12-base long, pyrimidine-rich sequence with an indispensible GTG motif (Beck and Lipps, 2007).

A highly related protein, Rep, also of the prim-pol family has both helicase and AEP primase-polymerase domains; Rep has been identified and characterised from the pIT3 plasmid of *Sulfolobus solfataricus* (Prato et al., 2008). The replicative N-terminal domain of this protein, termed Rep245, also harbours DNA polymerase activity, as well as dNTP and rNTP primase activities. Unlike ORF904 though, this archaeal plasmid primase domain has 3' terminal deoxynucleotide transferase activities (Prato et al., 2008). It remains to be seen whether this terminal transferase activity is retained for the full-length Rep protein.

Bacteria, like archaea, often harbour extrachromosomal plasmid DNA. Two decades ago, a Rep protein from the colicin E2 (ColE2) plasmid was found to have DNA primase activity (Takechi and Itoh, 1995; Takechi et al., 1995). 10 years later, it was shown that this primase was a member of the AEP family, distantly related to the prim-pols of ORF904 (pRN1) and Rep (pIT3) (L. M. Iyer et al., 2005). This protein, like

ORF904 but unlike other AEPs, shows sequence specificity to a defined origin sequence in its primase activity (Takechi and Itoh, 1995).

A protein with an AEP primase-polymerase domain has also been identified in the bacterial plasmid of *Bacillus cereus* RSF1010 (Samuels et al., 2009). This protein was identified through a homology search for eukaryotic MCMs and is formed of an AEP primase-polymerase domain fused with a helicase domain at its C-terminus (Samuels et al., 2009; Sanchez-Berrondo et al., 2012).

A third plasmid-encoded primase family has recently been described that is formed of subunits analogous to the PriS catalytic subunit and PriL regulatory subunit of the replicative AEPs in a single protein product (Gill et al., 2014). It is from the pTN2 plasmid from *Thermococcus nautilus* and is denoted PolpTN2 (Gill et al., 2014). The only other member of this family to be identified is from the pP12-1 plasmid of *Pyroccocus sp.12-1* (Soler et al., 2010). This primase-polymerase will only prime and extend from a primer using dNTPs, it will not utilise rNTPs (Gill et al., 2014). In addition to these activities, PolpTN2 has terminal transferase activity, which is greatly enhanced upon the removal of the PriL-like portion of the protein (Gill et al., 2014). This removal of the PriL-like portion of PolpTN2 also confers reverse transcriptase activity to the primase (Gill et al., 2014).

The zinc-binding motif of primases is present across every AEP primase described with the exceptions of PolpTN2 and Rep(pIT3) (Gill et al., 2014; Prato et al., 2008). Even the ORF904 primase of the pRN1 plasmid contains a zinc-binding motif and has a conserved active core in common with replicative AEPs (Lipps et al., 2004). However, the zinc-binding motif of ORF904 is in a distinct position relative to the zinc-binding motif of the replicative primases (See Figure 1.21) (Lipps et al., 2004). This hints at the two primase families having an evolutionary ancestor that had no zinc finger. Perhaps these plasmid gene products, PolpTN2 and Rep(pIT3), represent evolutionarily ancestral AEP primases.

Kinetoplastids are a group of single-celled protozoa characterised by the presence of kinetoplasts; a kinetoplast is a network of circular DNA, both maxicircles (20-40 kb) and minicircles (0.5-1 kb) found inside a single, large mitochondrion (reviewed in (Shapiro and Englund, 1995)). Until recently, the kinetoplastid replication machinery could not be

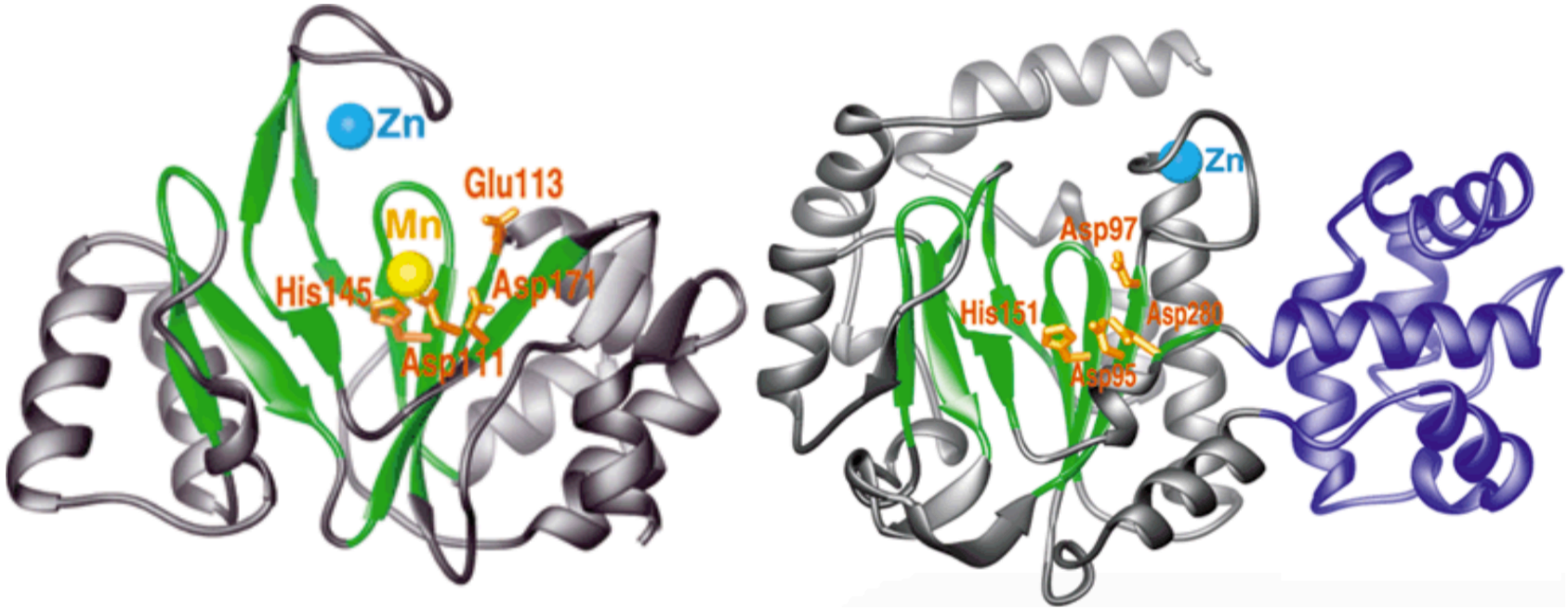


Figure 1.21. Difference in location of zinc-binding motif of replicative AEPs and ORF904 of the plasmid pRN1.
The zinc-binding motif of replicative AEPs (left) is in a distinct position relative to the conserved AEP core (green) to ORF904 (right), suggesting that the AEPs potentially evolved from a zinc-independent plasmid. This figure is adapted from Lipps et al. 2004.

reconstituted because it was not capable of *de novo* nucleotide chain synthesis, but the primase for this system had not been identified. Two such primases have since been characterised found in kinetoplastid mitochondria, PRI1 and PRI2, responsible for maxicircle and minicircle replication initiation, respectively (Hines and Ray, 2011; 2010). These primases are capable of producing primers of up to 70-80 nucleotides in length. PRI1 and PRI2 are from the nucleo-cytoplasmic large DNA virus (NCLDV)-herpesvirus clade of primases, which suggests they are an archaic primase acquired from a viral source (L. M. Iyer et al., 2005).

1.7.2.5. Viral AEPs

The UL5-UL8-UL52 heterotrimeric primase-helicase complex in herpes viruses is perhaps the most well characterised of the viral primases (Crute and Lehman, 1991). The UL5 protein is responsible for the helicase activities of the complex (Gorbalenya et al., 1989) and the UL52 protein is an AEP responsible for the priming activities of the complex (Crute and Lehman, 1991; Klinedinst and Challberg, 1994). Whilst UL8 is not a requirement for either the core primase or helicase activities, it is predicted to be required for the efficient utilisation of the primers by polymerase (Marsden et al., 1997; Sherman et al., 1992). A recent bioinformatics report has suggested that UL8 is distantly related to the B family of polymerases but has lost its active site and has lost most of its DNA binding motifs (Kazlauskas and Venclovas, 2014). Where most primases have a zinc-binding motif in their catalytic domains (Augustin et al., 2001; Lao-Sirieix et al., 2005; Lipps et al., 2004), UL52 has a strand-rich zinc finger domain that is separate at its C-terminus that is absolutely required for the protein's activity *in vivo* (Biswas and S. K. Weller, 1999). Similar to the pol α primase in eukaryotes, the UL52 primase produces ribonucleotide primers of around 8-12 nucleotides in length (Crute and Lehman, 1991).

The Lef-1-like primases of baculoviruses represent a family of AEPs that are closely related to the replicative primases of eukaryotes and archaea (L. M. Iyer et al., 2005). However, they are capable of producing RNA primers that extend up to several kilobases in length (Mikhailov and Rohrmann, 2002), suggesting an RNA polymerase activity for these primases. This is a trait shared by the D5-like proteins of poxvirus NCLDVs (De Silva et al., 2009). D5 has been characterised from the vaccinia virus and is conserved across all poxviruses (De Silva et al., 2009). D5, unlike Lef-1, is formed of an AEP polymerase domain and a helicase domain. (De Silva et al., 2009).

Another family of AEPs, the gp43-like proteins of corynephage BFK20, does not share this incorporation of rNTPs in viral AEPs (Halgasova et al., 2012). gp43 is related to the prim-pol family of proteins, including ORF904 and Rep(pIT3), and has both primase and polymerase activities but will only incorporate dNTPs, not rNTPs (Halgasova et al., 2012). Unlike prim-pols ORF904 and the Rep protein of ColE2, gp43 does not show any apparent sequence specificity (Halgasova et al., 2012). This gp43-like protein has been identified in at least 13 mycobacteriophages and ReqiPine5, a *Rhodococcus* phage (Halgasova et al., 2012).

Taken together, these studies establish that AEPs fulfil a wide range of tasks in archaea, bacteria and viruses including priming, polymerase, lesion bypass, terminal transferase, gap filling and strand displacement activities. This plethora of activities illustrates that AEPs function in both DNA repair mechanisms and DNA damage tolerance during replication, in addition to the known role in priming DNA for replication. This suggests that during the evolution of eukaryotes with larger genomes, cells developed more complex mechanisms using a variety of specialised polymerases for DNA damage repair and tolerance, in particular the X and Y families of polymerases that remain absent from most archaeal and prokaryotic species. As such, the pol α primase is only required for the DNA-dependent RNA priming of DNA in eukaryotes.

1.8. Discovery of Eukaryotic Primase-Polymerase

Until recently, the only AEP identified in higher eukaryotes was the pol α-associated primase small subunit PriS. An *in silico* analysis of the AEP superfamily identified the existence of an additional uncharacterised DNA primase in eukaryotes called Primase-Polymerase, or PrimPol (alternative names CCDC111, FLJ33167, EukPrim2 and hPrimPol1) (L. M. Iyer et al., 2005). This novel protein belongs to the NCLDV-herpesvirus clade of AEPs, which includes the mitochrondrial primases of kinetoplastids, UL52 of the herpesviruses and D5 of the poxviruses (Section 1.7.2.5)(L. M. Iyer et al., 2005). PrimPol is present across a broad range of unicellular and multicellular eukaryotes across animals, plants, protists and fungi (Figure 1.22). It is, however, notably absent in a number of eukaryotes, including *Caenohabditis elegans* and *Drosophila melanogaster*, and has only been identified in a small handful of fungal species. This distribution suggests that PrimPol was acquired from a viral source

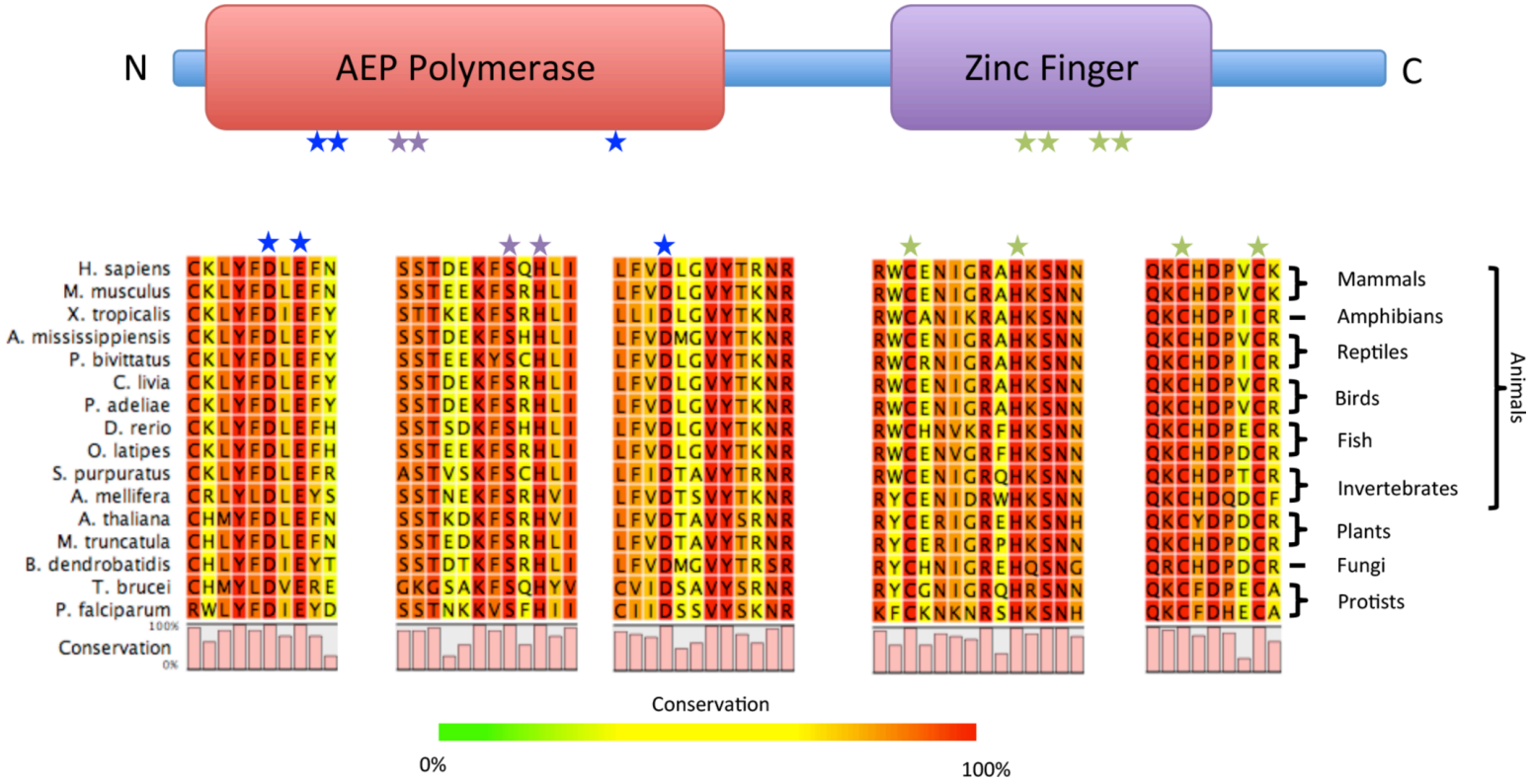


Figure 1.22. Conserved motifs of PrimPol.
PrimPol is conserved across a range of eukaryotes from protists to humans. PrimPol consists of an N-terminal AEP polymerase domain and C-terminal putative zinc-finger domain. All PrimPol family members share conserved motifs; these motifs include the invariant DhE of AEP motif I (blue stars), SxH of AEP motif II (purple stars), hD of motif III (blue stars) and the -C-H-C-C- zinc coordinating amino acids of the zinc finger (green stars).

through horizontal gene transfer and subsequently lost independently on a range of occasions (L. M. Iyer et al., 2005).

PrimPol is composed of two discernable domains, an AEP polymerase domain at its N-terminus and a putative UL52-like zinc finger at its C-terminus (Figure 1.22). The AEP polymerase consists of the three canonical motifs of an AEP, motif I shares the consensus sequence LYFDLE, motif II shares the consensus sequence SxH and motif III the consensus sequence VD. The only other AEP identified so far with a glutamate residue in the final position of motif I (i.e. hhhDxE) is ORF904 (Figure 1.17). The zinc finger of PrimPol is a CHCC-type zinc finger and is conserved across the NCLDV-herpesvirus clade of AEPs (L. M. Iyer et al., 2005). Recent characterisation of PrimPol by a number of groups has shown that this protein has both DNA-dependent primase and polymerase activity, hence the designation of the name PrimPol (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Keen et al., 2014; Mourón et al., 2013; Rudd et al., 2014; 2013; Wan et al., 2013).

PrimPol is required for the efficient progression of the replication fork during the replication of chromosomal DNA (Bianchi et al., 2013; Mourón et al., 2013; Wan et al., 2013). It also plays a role in the maintenance of mitochondrial DNA replication (Garcia-Gomez et al., 2013). This is consistent with the finding that, upon fractionation of HeLa cells, PrimPol is distributed between the cytosol (47%), nucleus (19%) and mitochondria (34%) (Garcia-Gomez et al., 2013).

This thesis aims to investigate the biochemical and biophysical properties of PrimPol and how this contributes to the roles it plays in both standard and perturbed replication of chromosomal and mitochondrial DNA. We started this study by probing the domain architecture of PrimPol (Chapter 3) and investigating which parts of the protein are required for the different catalytic and DNA binding activities associated with PrimPol. The enzymatic characterisation of PrimPol is also explored in Chapter 4; this chapter examines both its kinetic parameters relative to other primases and polymerases, as well as its lesion bypass proficiencies in the presence and absence of the zinc finger of PrimPol. Chapter 5 elucidates the biochemical characteristics of a mutation in PrimPol that has been reported to have clinical relevance (Zhao et al., 2013). The final chapter (Chapter 6) introduces the work towards researching PrimPol's binding partners *in vivo*. These chapters will also describe the additional background studies that were carried

out in the Doherty lab in parallel with this study that complement this work. Each chapter will be briefly introduced by the experiments that preceded and informed the studies described in the chapter to supplement my introduction to PrimPol here.

CHAPTER 2

Materials and Methods

2.1. Molecular Biology Methods

2.1.1. Production of Competent DH5α E. coli

DH5α cells were plated on LB agar plates (LB medium solidified with 1.5% (w/v) agar; LB = lysogeny broth (1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl, pH 7)). 3 mL of LB was then inoculated with a single colony of DH5α cells selected from this plate and grown overnight at 37°C with shaking at 180 rpm. This culture was then diluted in 250 mL super optimal broth (SOB) medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$ and 10 mM $MgCl_2$) and placed in an 18°C incubator until an optical density at 600 nm (OD_{600}) of 0.4. Cells were cooled for 10 minutes on ice and then collected by centrifugation (4,000 rpm, 10 minutes, 4°C, Sorvall Legend RT, 75006445 rotor). The cells were resuspended in transformation buffer (100 mM PIPES (pH 6.7), 15 mM $CaCl_2$, 250 mM KCl, 55 mM $MgCl_2$) and incubated for 10-30 minutes on ice. Cells were then pelleted again and resuspended in 15 mL transformation buffer supplemented with 7% (v/v) dimethyl sulphoxide (DMSO). Cells were incubated on ice for 10 minutes before aliquoting into 50-100 µL aliquots and frozen in liquid nitrogen and stored at -80°C.

2.1.2. Transformation of Competent E. coli

1 µL of plasmid DNA from a Miniprep (~50-100 ng)(Section 2.1.3) or 5 µL of a ligation reaction (Section 2.1.7) was added to 50 µL chemically competent *E. coli* cells. The *E. coli* and DNA mix was incubated on ice for 30 minutes prior to a 35 second heat shock at 42°C before returning to ice for 2 minutes. 300 µL of SOB was added before incubation at 37°C for 1 hour. 100 µL of this cell growth was plated on a pre-warmed LB agar plate containing appropriate antibiotics and incubated overnight at 37°C to allow colony growth. Kanamycin was used at 34 µg mL^{-1} final concentration.

2.1.3. Plasmid DNA Amplification and Purification

Single E. coli colonies that had been transfected with plasmid DNA (Section 2.1.2.) were inoculated into 5 mL of LB medium containing the appropriate antibiotic and incubated at 37°C overnight with shaking at 180 rpm. Following overnight incubation, the plasmid DNA was purified using the QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Plasmid DNA was eluted in 50 µL and stored at -20°C. The yield of the plasmid DNA was determined using an ND-1000 NanoDrop spectrophotometer (Thermo Scientific) at a wavelength of 260 nm.

2.1.4. Agarose Gel Electrophoresis of DNA

DNA was resolved on a 1% (w/v) agarose gel containing TAE (200 mM Tris, 100 mM Acetic Acid, 5 mM EDTA) and ~0.3 µg mL^{-1} ethidium bromide. DNA samples were supplemented DNA loading buffer (5% (v/v) glycerol, 3.3 mM Tris pH 8, 0.04% bromophenol blue) prior to loading and were loaded alongside a 1 kb DNA ladder (New England Biolabs) or 100 bp DNA ladder (New England Biolabs) as appropriate. Gels were electrophoresed at 100-120 V for ~20 minutes in TAE buffer. DNA was visualised using a UV illuminator (Syngene InGenius Gel Analysis System) and images were analysed using GeneSnap (Syngene)

2.1.5. Polymerase Chain Reaction (PCR)

PCR reactions of 50 µL were assembled in 200 µL reaction tubes using 1 µM of each of a reverse and forward primer, ~5 ng of template plasmid, 1 mM dNTPs, 10 mM $MgCl_2$, 20% DMSO and 1 U Phusion high-fidelity DNA polymerase (New England Biolabs). The primers were designed to have a typical T_m ~55-65°C and are flanked by an appropriate restriction digestion site and a lead 'anchor' sequence for efficient digestion; the nucleotide sequences of primers used can be found in Table 2.1. The PCR conditions used for each reaction (i.e. temperatures, times and cycles) can be found in Table 2.2. Reactions were carried out using a Professional Trio Thermocycler (Biometra). PCR products were confirmed by agarose gel electrophoresis (Section 2.1.4), loaded alongside a no-polymerase control.

Upon confirmation of DNA amplification, PCR products were electrophoresed on an agarose gel (Section 2.1.4) and detected using a UV transilluminator (UVP). Bands were excised from the gel and purified using the QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's instructions.

2.1.6. Restriction Digestion

Typically 20 µL of purified plasmid DNA or 30 µL of PCR product was digested in a 50 µL reaction containing 10 U of each of the appropriate double-digestion restriction enzymes (New England Biolabs), 100 µg mL^{-1} bovine serum albumin (BSA), and the appropriate restriction buffers. Reactions were incubated at 37°C for 3-4 hours. Purified plasmid DNA was supplemented with Antarctic phosphatase buffer (50 mM Bis-Tris-

#	Primer	Sequence (5'-3')
1	Human PrimPol 1 Forward	GTTTCTTCATATGAATAGAAAATGGGAAGCAAAACTG
2	Human PrimPol 24 Forward	GTTTCTTCATATGCCGTTGTCCTCAGTGTATAGACC
3	Human PrimPol 39 Forward	GTTTCTTCATATGTCCATCTGGAGACTATTTCATCGAC
4	Human PrimPol 372 Forward	GTTTCTTCATATGCAGTGTTCTCCCTATCCTGAAGTTG
5	Human PrimPol 480 Forward	GTTTCTTCATATGACAACAGATGAAGCAGATGAAAC
6	Human PrimPol 276 Reverse	CAAAGAAGCGGCCGCCTAATGCTTCTCTCCCATGTTATTCTTCAC
7	Human PrimPol 354 Reverse	CAAAGAAGCGGCCGCCTATTTTTGTTTATTCTGAGATGGCTCAC
8	Human PrimPol 560 Reverse	CAAAGAAGCGGCCGCTTACTCTTGTAATACTTCTATAATTAGTTC
9	Human PrimPol K300stop Forward	CATCAAAAATTGGATAGCGTGTGGCTTTGG
10	Human PrimPol K300stop Reverse	CCAAAGCCACACGCTATCCAATTTTTGATG
11	Human PrimPol R488stop Forward	GAAACTTAGAGCAATGAAACCCAGAATCC
12	Human PrimPol R488stop Reverse	CATTGCTCTAAGTTTCATCTGCTTCATC
13	Human PrimPol C419A/H426A Forward	GTAAATATCGGTGGGCTGAAAACATTGGAAGAGCCGCTAAGAGTAATAATATAATG
14	Human PrimPol C419A/H426A Reverse	CTCTTAGCGGCTCTTCCAATGTTTTCAGCCCACCGATATTTACAAATATCATAAACCAGTAATTC
15	Human PrimPol Y89D Forward	GAATTTTGGTTTGACTATAAATCCAGAAAAAATCTCTTACACTGCTATG
16	Human PrimPol Y89D Reverse	GATTTATAGTCAAACCAAAATTCAGCATAGGTTGTCACAAG
17	Human PrimPol Y89F Forward	GAATTTTGGTTTTTCTATAAATCCAGAAAAAATCTCTTACACTGCTATG
18	Human PrimPol Y89F Reverse	GATTTATAGAAAAACCAAAATTCAGCATAGGTTGTCACAAG
19	Human PrimPol Y89S Forward	GAATTTTGGTTTTCATATAAATCCAGAAAAAATCTCTTACACTGCTATG
20	Human PrimPol Y89S Reverse	GATTTATATGAAAACCAAAATTCAGCATAGGTTGTCACAAG
21	Mouse PrimPol 24 Forward	GTTTCTTCATATGAGACCAAGACTGGCCAAGCCAGAAG
22	Mouse PrimPol 39 Forward	GTTTCTTCATATGTCCATCTGGAGACTATTTCATCG
23	Mouse PrimPol 338 Reverse	CAAAGAAGCGGCCGCCTAAGCCCGCTTTCGTTTAGTCTGAGATGGG
24	Mouse PrimPol S315stop Forward	CTGGTGTGAAATGTCAGATTCTCAG
25	Mouse PrimPol S315stop Reverse	GACATTTCACACCAGGGAAGATAGG
26	Xenopus PrimPol 1 Forward	GTTTCTTGGATCCATGAAAGCATGTGGAGGACTGGCAGATTATTATAG
27	Xenopus PrimPol 699 Reverse	GACATGTAGCAACCCTGACTATCGAACAAGCAAATAGCTCGAGAAGAAAC

Table 2.1. Primers used in PCR and site directed mutagenesis to produce expression vectors. The primers listed here are used in the construction of expression vectors listed in Table 2.3.

#	Initialisation	Denaturation	Annealing	Extension	Cycles	Finalisation	Hold
1	98°C (1:00)	95°C (0:35)	56°C (0:35)	68°C (1:10)	30	68°C (3:30)	4°C
2	95°C (1:00)	95°C (0:30)	55°C (0:30)	70°C (7:30)	22	70°C (15:00)	4°C
3	95°C (1:00)	95°C (0:30)	55°C (0:30)	60°C (7:30)	21	60°C (15:00)	4°C
4	98°C (1:00)	95°C (0:35)	56°C (0:35)	72°C (3:00)	30	72°C (9:00)	4°C

Table 2.2. PCR methods used in the generation of novel expression plasmids.
The PCR conditions used to construct the expression vectors in Table 2.3 are listed here. The initialisation, denaturation, annealing, extension and finalisation temperatures are listed with the associated times in brackets as (minutes:seconds). The cycles indicate cycling between denaturation, annealing and extension steps with the number of cycles listed.

Propane-HCl pH 6, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$) and treated with Antarctic phosphatase (New England Biolabs) for 30 minutes to prevent plasmid recircularisation. The restriction products were purified using the QIAquick PCR Purification kit (Qiagen) according to the manufacturer's instructions.

2.1.7. Ligation of DNA

Restriction digested plasmid DNA and PCR products were ligated using T4 DNA ligase (New England Biolabs). The concentration of digested DNA was determined using a ND-1000 NanoDrop spectrophotometer (Thermo Scientific) at a wavelength of 260 nm. Ligation reactions were set up at a ratio (n) of 6:1 or 3:1 of Insert (I) to Vector (V) using the following equation:

$$I_{(ng)} = \frac{n \; x \; I_{(bp)} \; x \; V_{(ng)}}{V_{(bp)}}$$

Reactions were supplemented with 200 units T4 DNA ligase, T4 DNA ligase buffer (50 mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT) and incubated at room temperature for 1 hour. 5 µL of reaction was transformed into DH5α (Section 2.1.2). The remainder of the ligation was incubated overnight at 12°C and transformed again if required.

Ligations were typically set up twice and the second reaction was subjected to heat inactivation after 1 hour at 95°C for 10 minutes. This reaction was then subject to restriction digestion by a restriction enzyme that was present in the polylinker of the vector that had been excised prior to ligation but not present in the PCR product to ensure any recircularised vector is linearised. This reaction was incubated at room temperature for 1 hour and subsequently 5 µL of reaction was transformed into DH5α (Section 2.1.2).

Colonies were screened by restriction digest (Section 2.1.6) and sequences confirmed by DNA sequencing (Section 2.1.9).

2.1.8. Site-Directed Mutagenesis

Site-directed mutagenesis primers were designed using either the QuickChange protocol (Stratagene), using complementary primer pairs, or a modified version using

partially overlapping primers to reduce primer dimerisation (Zheng et al., 2004). Primers for site-directed mutagenesis can be found in Table 2.1. Site-directed mutagenesis reaction reagents are the same as those in Section 2.1.5. and the reaction conditions for PCR can be found in Table 2.2. Products were screened by agarose gel electrophoresis (Section 2.1.4) alongside a no polymerase control. Successfully amplified products were digested using 10 U DpnI (New England Biolabs) in a 30 µL reaction for ~3 hours at 37°C. 3 µL of reaction products were transformed into DH5α (Section 2.1.2). Positive site-directed mutagenesis clones were confirmed by DNA sequencing (Section 2.1.9).

2.1.9. Sequencing of DNA products

Sequencing of plasmid DNA was performed by GATC biotech using universal primers or gene-specific primers. The sequencing chromatogram was read using 4Peaks (Mekentosj).

2.2. Purification of Recombinant Proteins

2.2.1. Preparation of Chemically Competent E. coli strains

Chemically competent BL21, Tuner and Rosetta *E. coli* strains were prepared in the same manner as DH5α (Section 2.1.1). Chemically competent SHuffle were also prepared in the same manner with the exception that they were not grown at a temperature above 30°C.

2.2.2. Strain Optimisation of Protein Expression

Chemically competent *E. coli* expression strains were transformed (Section 2.1.2) with expression plasmids (Table 2.3). Colonies from each expression strain were inoculated into 5 mL of LB medium supplemented with 34 µg mL^{-1} kanamycin and incubated overnight at 37°C (30°C for SHuffle cells) with shaking at 180 rpm. 1 mL of overnight growth was added to 50 mL of LB supplemented with 34 µg mL^{-1} kanamycin and allowed to continue to grow. Upon reaching an OD_{600} of 0.4, as determined by spectrophotometry on a Cary 50 Bio spectrophotometer (Varian), growths were rapidly cooled on ice for 10 minutes. Expression was subsequently induced through the addition of 400 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and growths were split into incubators at temperatures ranging from 16-30°C for 3-16 hours.

#	Gene Product	Vector	Source	Primers	PCR Method
H1	Wild-Type Human PrimPol	pET28a	Dr. Julie Bianchi	N/A	N/A
H2	Human PrimPol 1-354	pET28a	PCR, plasmid H1	1, 7	1
H3	Human PrimPol 24-354	pET28a	PCR, plasmid H1	2, 7	1
H4	Human PrimPol 39-354	pET28a	PCR, plasmid H1	3, 7	1
H5	Human PrimPol 1-276	pET28a	PCR, plasmid H1	1, 6	1
H6	Human PrimPol 39-276	pET28a	PCR, plasmid H1	3, 6	1
H7	Human PrimPol 39-299	pET28a	SDM, plasmid H4	9, 10	2
H8	Human PrimPol 1-487	pET28a	SDM, plasmid H1	11, 12	2
H9	Human PrimPol 372-560	pET28a	PCR, plasmid H1	4, 8	1
H10	Human PrimPol 480-560	pET28a	PCR, plasmid H1	5, 8	1
H11	Human PrimPol C418A/H425A (ZF-KO)	pET28a	SDM, plasmid H1	13, 14	2
H12	Human PrimPol 372-560 ZF-KO	pET28a	SDM, plasmid H9	13, 14	2
H13	Human PrimPol Y89D	pET28a	SDM, plasmid H1	15, 16	2
H14	Human PrimPol 1-354 Y89D	pET28a	SDM, plasmid H2	15, 16	2
H15	Human PrimPol Y89F	pET28a	SDM, plasmid H1	17, 18	2
H16	Human PrimPol Y89S	pET28a	SDM, plasmid H1	19, 20	2
H17	Human PrimPol D514R,D518R,D519R,D524A (RBM-A-KO)	pET28a	Dr. Nigel Brissett	N/A	N/A
H18	Human PrimPol 480-546	pET28a	Dr. Nigel Brissett	N/A	N/A
H19	Human PrimPol 480-546/RBM-A-KO	pET28a	Dr. Nigel Brissett	N/A	N/A
H20	Human PrimPol D514R, D518R, D519R, D551R, I554A, I555A (RBM-A-KO/RBM-B-KO)	pET28a	Dr. Nigel Brissett	N/A	N/A
M1	Mouse PrimPol 1-338	pET28a	Dr. Laura Bailey	N/A	N/A
M2	Mouse PrimPol 24-338	pET28a	PCR, plasmid M1	21, 23	1
M3	Mouse PrimPol 39-338	pET28a	PCR, plasmid M1	22, 23	1
M4	Mouse PrimPol 1-314	pET28a	SDM, plasmid M1	24, 25	3
X1	Xenopus PrimPol Full Length	pET28a	PCR, Genscript Synthesis	26, 27	4
X2	Xenopus PrimPol 1-334	pET28a	Dr. Stanislaw Jozwiakowski	N/A	N/A
X3	Xenopus PrimPol 1-472	pET28a	Dr. Stanislaw Jozwiakowski	N/A	N/A

Table 2.3. Plasmid expression vectors and their associated gene products.
All gene products are located within the polylinker of the pET28a plasmid for expresion. The source is given with polymerase chain reaction (PCR) and site-directed mutagenesis (SDM) products indicated and the plasmid that was used as a template. Those plasmids that were donated by other scientists are also indicated. Plasmid X1 was subcloned from a gene synthesised by Genscript. The PCR/SDM primers used are the primer numbers (#) from Table 2.1. The PCR method used are the method numbers (#) listed in Table 2.2.

1.5 mL samples were taken from each growth. Cells were resuspended in lysis buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.5, 30 mM imidazole, 10% (v/v) glycerol, 17 µg mL^{-1} phenylmethanesulfonylfluoride (PMSF), 34 µg mL^{-1} benzamidine, 0.1% IGEPAL). Samples were sonicated using a Vibra-Cell sonicator (Sonics) for 10 seconds at 30% amplitude and centrifuged in a benchtop centrifuge at 13,000 rpm for 1 hour at 4°C to clear insoluble cell debris from the lysate. 0.5 µL of the soluble fraction was diluted into buffer then subjected to sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Section 2.3.1) to analyse protein expression in different strains, at different induction temperatures, at different induction lengths.

2.2.3. Recombinant Protein Expression and Preparation

Following strain optimisation if required, colonies from protein expression strains of *E. coli* are inoculated into 5 mL of LB medium supplemented with 34 µg mL^{-1} kanamycin and incubated overnight at 37°C (30°C for SHuffle cells). 3 mL of overnight growth was then added to a large volume (1-6 litres) of LB medium or terrific broth (1.2% tryptone, 2.4% yeast extract, 72 mM K_2HPO_4, and 1% (v/v) glycerol) (TB) and grown at 37°C (30°C for shuffle cells) with shaking at 180 rpm. Upon reaching an optical density at an OD_{600} of 0.4 for growth in LB or 0.6 for growth in TB, growths were cooled at for 30 minutes at 4°C and subsequently induced through the addition of 400 µM IPTG and allowed to continue to grow. The growth and induction conditions for each of the proteins in this thesis can be found in Table 2.4.

Cells were collected by centrifugation (4,000 rpm, 15 minutes, 4°C, Sorvall Evolution, SLC-6000 rotor) and resuspended in lysis buffer (15 mL per litre of culture grown). To this resuspension, 1 mg mL^{-1} of chicken egg lysozyme was added and this stirred for 30-45 minutes. Lysate was then sonicated using a Vibra-Cell sonicator (Sonics) at 30% amplitude for 6 second pulses, with a 10 second rest between pulses, for a total of 6 minutes sonication time. Insoluble cell debris was cleared by centrifugation (18,000 rpm, 1 hour, 4°C, JA 25-50 rotor). The soluble fraction was then passed through a 0.45 µM filter (Millipore) and kept on ice for protein purification by chromatography.

2.2.4. Affinity Chromatography of Recombinant Proteins

Table 2.5 outlines the purification strategies for the different proteins purified in the scope of this work. Immobilised metal affinity chromatography (IMAC) was used to purify all recombinant proteins fused with an N-terminal 6x-His tag. 5 mL or 25 mL of

nickel-nitrolotriacetic acid (Ni-NTA) agarose (Qiagen) was packed into chromatography columns. These Ni-NTA columns were then subjected to chromatography using an ÄKTAprime (GE Healthcare) or ÄKTAprime plus (GE Healthcare) system. All buffers and protein samples were kept on ice. Columns were pre-equilibrated using IMAC Buffer A (300 mM NaCl, 50 mM Tris-HCl pH 7.5, 30 mM imidazole, 10% (v/v) glycerol, 17 μg mL^{-1} phenylmethanesulfonylfluoride (PMSF), 34 μg mL^{-1} benzamidine, 2 mM β-mercaptoethanol). Cell lysates were loaded at a constant rate of 3 mL min^{-1}. Columns were then washed using buffer A until all non-specifically bound proteins were eluted as measured by absorbance at a wavelength of 280 nm (A_{280}) returning to basal levels. Columns were subjected to a wash using 5% IMAC Buffer B (as IMAC Buffer A, except 300 mM imidazole) to remove any weakly bound proteins. The bound proteins were then eluted using 100% buffer B. Samples were taken from the insoluble fraction, load, flow-through, wash, 5% buffer B wash and 100% B elution and subjected to SDS-PAGE analysis (see Section 2.3.1).

Some proteins (see Table 2.5) were subsequently purified through the use of a 5 mL HiTrap Heparin HP column (GE Healthcare) connected to an ÄKTAprime (GE Healthcare) or ÄKTAprime plus (GE Healthcare) system. Heparin-sepharose functions as a mixed-mode chromatography column as both an affinity column for DNA-binding proteins, as well as a cation exchanger. Columns were equilibrated using pre-chilled Heparin Buffer A (50 mM NaCl, 50 mM Tris-HCl pH 7.5, 10% (v/v) glycerol, 2 mM DTT). Proteins were diluted 1:10 into Heparin Buffer A to reduce the salt and imidazole concentrations and loaded onto the column at a flow rate of 2.5 mL min^{-1}. The column was then washed using Heparin Buffer A to elute any non-specifically bound proteins. Proteins were eluted by gradient elution up to 50% Heparin Buffer B (as Heparin Buffer A but 2 M NaCl). Protein-contaning fractions were determined by A_{280} level and samples of these fractions were subjected to SDS-PAGE analysis (see Section 2.3.1).

2.2.4. Ion-Exchange Chromatography Recombinant Proteins

Some recombinant proteins (see Table 2.5) were subjected to ion-exchange chromatography based on their predicted isoelectric point (pI) (see Section 2.5.4). 5 mL HiTrap Q Sepharose FF columns (GE Healthcare) were used as anion exchangers and 5 mL HiTrap SP Sepharose FF (GE Healthcare) were used as cation exchangers connected to an ÄKTAprime (GE Healthcare) or ÄKTAprime plus (GE Healthcare)

Protein	E. coli strain	Growth Medium	Induction Temperature and Length
Wild-Type Human PrimPol	SHuffle	LB	16°C overnight
Human PrimPol 1-354	BL21	TB	25°C overnight
Human PrimPol 24-354	BL21	TB	25°C overnight
Human PrimPol 39-354	BL21	TB	25°C overnight
Human PrimPol 1-276	BL21	LB	25°C overnight
Human PrimPol 39-276	BL21	LB	25°C overnight
Human PrimPol 39-299	BL21	LB	25°C overnight
Human PrimPol 1-487	SHuffle	LB	16°C overnight
Human PrimPol 372-560	BL21	LB	25°C overnight
Human PrimPol 480-560	BL21	TB	20°C overnight
Human PrimPol C418A/H425A (ZF-KO)	SHuffle	LB	16°C overnight
Human PrimPol 372-560 ZF-KO	BL21	LB	25°C overnight
Human PrimPol Y89D	SHuffle	LB	16°C overnight
Human PrimPol 1-354 Y89D	BL21	LB	25°C overnight
Human PrimPol Y89F	SHuffle	LB	16°C overnight
Human PrimPol Y89S	SHuffle	LB	16°C overnight
Human PrimPol RBM-A-KO	BL21	TB	20°C overnight
Human PrimPol 480-546	BL21	TB	20°C overnight
Human PrimPol 480-546/RBM-A-KO	BL21	TB	20°C overnight
Human PrimPol RBM-A-KO/RBM-B-KO	BL21	TB	20°C overnight
Mouse PrimPol 1-338	BL21	TB	20°C overnight
Mouse PrimPol 24-338	BL21	TB	22.5°C overnight
Mouse PrimPol 39-338	BL21	TB	25°C overnight
Mouse PrimPol 1-314	BL21	LB	20°C overnight
Xenopus PrimPol Full Length	SHuffle	TB	16°C overnight
Xenopus PrimPol 1-334	BL21	TB	25°C overnight
Xenopus PrimPol 1-472	SHuffle	LB	16°C overnight

Table 2.4. Growth conditions for recombinant PrimPol proteins.
The strain of *E. coli* used in the growth of recombinant proteins is listed along with the growth medium used. *E. coli* growths were induced at an OD_{600} of 0.4 for Lysogeny Broth (LB) or 0.6 for Terrific Broth (TB), following a period of rapid cooling. The induction temperature and length is listed.

Protein	Affinity		Ion Exchange		Size Exclusion
	Ni-NTA IMAC	Heparin	Q	SP	Gel Filtration
Wild-Type Human PrimPol	✓	✓			
Human PrimPol 1-354	✓	✓		✓	✓
Human PrimPol 24-354	✓	✓			✓
Human PrimPol 39-354	✓	✓			✓
Human PrimPol 1-276	✓				
Human PrimPol 39-276	✓				
Human PrimPol 39-299	✓				
Human PrimPol 1-487	✓	✓			
Human PrimPol 372-560	✓		✓		
Human PrimPol 480-560	✓		✓		✓
Human PrimPol C418A/H425A (ZF-KO)	✓	✓			
Human PrimPol 372-560 ZF-KO	✓		✓		
Human PrimPol Y89D	✓	✓			
Human PrimPol 1-354 Y89D	✓	✓			
Human PrimPol Y89F	✓	✓			
Human PrimPol Y89S	✓	✓			
Human PrimPol RBM-A-KO	✓		✓		✓
Human PrimPol 480-546	✓		✓		✓
Human PrimPol 480-546/RBM-A-KO	✓		✓		✓
Human PrimPol RBM-A-KO/RBM-B-KO	✓		✓		✓
Mouse PrimPol 1-338	✓	✓		✓	✓
Mouse PrimPol 24-338	✓	✓		✓	✓
Mouse PrimPol 39-338	✓	✓		✓	✓
Mouse PrimPol 1-314	✓	✓			
Xenopus PrimPol Full Length	✓	✓			
Xenopus PrimPol 1-334	✓	✓			✓
Xenopus PrimPol 1-472	✓				

Table 2.5. Purification strategies for recombinant PrimPol proteins.
The chromatography columns used in the purification of recombinant proteins are listed.

system. Ion-exchange chromatography was carried out in the same manner as chromatography using a HiTrap Heparin HP column (see Section 2.2.3).

2.2.5. Size-exclusion chromatography

The final step of purification for some of the proteins (see Table 2.5) was size-exclusion chromatography (SEC) using either an S75 or S200 10/300 GL gel-filtration column (GE Healthcare), depending on protein size. An ÄKTApurifier 10 or ÄKTAFPLC system was used for SEC. Columns were pre-equilibrated with SEC Buffer (typically 50 mM Tris-HCl (pH7.5), 300 mM NaCl, 10% (v/v) glycerol and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP)) that had been sterile-filtered using a 0.2 µm pore size vacuum filtration system (Nalgene). Protein samples were concentrated to less than ~4 mL using a Vivaspin sample concentrator (GE Healthcare) with an appropriate molecular weight cut off. These protein samples were loaded onto a 5 mL loop that was pre-filled with SEC Buffer. The SEC columns were run at a flow rate of 2.2 mL min^{-1} and fractions were collected following 100 mL of flow-through (void volume). Protein-containing fractions were determined by A_{280} level and samples of these fractions were subjected to SDS-PAGE analysis (see Section 2.3.1).

2.2.6. Removal of 6x-Histidine Tag

For some proteins that were to be screened for crystallisation, the hexahistidine tag at the N-terminal was removed following purification. The histidine tag linker in pET28a includes a thrombin cleavage site. Thrombin was used to cleave the histidine tag at a concentration of 1 unit per mg purified protein for 2 hours at room temperature. The thrombin and cleaved His-tag was then separated from the purified protein by chromatography using a 1 mL HisTrap HP column (GE Healthcare) connected in series to a HiTrap Benzamidine FF column (GE Healthcare). The cleaved protein sample was compared with uncleaved protein by SDS-PAGE (see Section 2.3.1).

2.2.7. Storage of Recombinant Proteins

Proteins-containing fractions were pooled and concentrated to a volume of less than 5 mL using a Vivaspin sample concentrator (GE Healthcare) with an appropriate molecular weight cut off. The protein concentration was determined by spectrophotometry using either an ND-1000 NanoDrop (Thermo Scientific) or Cary 50 Bio (Varian) spectrophotometer at a wavelength of 280 nm. If not present in the purification buffer, samples were supplemented with 10% (v/v) glycerol and aliquoted

into small volumes (typically 200 µL). Aliquots were snap frozen in liquid nitrogen and stored at -80°C.

2.3. Biochemistry Methods

2.3.1. SDS Polyacrylamide Gel Electrophoresis

Protein samples were resolved using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE gels were prepared in 1 mm Novex Gel cassettes (Invitrogen) and consisted a stacking gel layer cast over resolving gel (Sambrook and Russell, 2006). The resolving gel was typically 12% polyacrylamide (40% (v/v) acrylamide/bisacrylamide (37.5:1) 30% mix (National Diagnostics), 375 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.1% ammonium persulphate (APS), 0.04% N,N,N',N'-Tetramethylethylenediamine (TEMED)). The resolving gel was allowed to set under a layer of isopropanol, upon setting the isopropanol was discarded and topped with a layer of stacking gel of 5% polyacrylamide (17% (v/v) acylamide/bisacrylamide (37.5:1) 30% mix (National Diagnostics, 125 mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.04% (v/v) TEMED) before a well comb was added and allowed to set. The XCell SureLock Mini-Cell Electrophoresis System (Invitrogen) was used for SDS-PAGE. Samples were prepared through the addition of Laemmli sample buffer (2% (w/v) SDS, 10% v/v β-mercaptoethanol, 20% (v/v) glycerol, 0.02% bromophenol blue, 125 mM Tris-HCl (pH 6.5)) (Laemmli, 1970), and were boiled for 10 minutes at 95°C. Protein samples were run alongside a Precision Plus (Bio-Rad) or Precision Plus Dual Marker (Bio-Rad) for Coomassie staining (Section 2.3.1.1) or western blotting (Section 2.3.1.2) respectively. Samples were electrophoresed at 150 V until the bromophenol blue dye reached the bottom of the gel (typically ~90 minutes).

2.3.1.1. Coomassie Staining

SDS polyacrylamide gels were stained in Coomassie blue solution (50% (v/v) methanol, 10% (v/v) acetic acid, 0.5% (w/v) Coomassie brilliant blue) for ~10 minutes with shaking. Coomassie blue solution was then removed and gel was gently washed with water before the addition of destaining solution (10-20% (v/v) methanol, 10% (v/v) acetic acid) and returning to the rocker for 2-16 hours until sufficient decolouration of the background dye so that protein bands were properly visible. Gels were then stored in water with 1% (v/v) glycerol and air-dried using the GelAir Drying System (Bio-Rad).

2.3.1.2 Western Blotting

Samples to be analysed by western blot were initially resolved by SDS-PAGE (Section 2.3.1). The SDS polyacrylamide gel was equilibrated with transfer buffer (20mM Tris, 50mM glycine, 10% (v/v) methanol). Concurrently, a polyvinyladine fluoride (PVDF) membrane (Millipore) was activated with methanol. Proteins were transferred to the PVDF membrane using the XCell II Blot Module (Invitrogen) according to the manufacturer's instructions. Transfer was performed at 25 V for 60 minutes in transfer buffer. The membrane was then blocked with blocking buffer (Tris-buffered saline (TBS; 280 mM NaCl, 20 mM Tris) supplemented with 0.05% (v/v) Tween 20 and 5% (w/v) non-fat dried milk (Marvel)) for at least one hour at room temperature. The primary antibody was diluted into 2.5 mL fresh blocking buffer and added to the membrane and incubated on a roller at 4°C overnight. The membrane was then washed three times in TBS supplemented with 0.05% (v/v) Tween for ~10 minutes per wash. For western blots that required a secondary antibody, secondary antibody was added in 2.5 mL blocking buffer and incubated at room temperature for 1 hour on a roller. The membrane was then washed three more times. Following antibody exposure, chemiluminescent detection with Amersham ECL Western Blotting Detection reagent (GE Healthcare) was performed according to the manufacturer's instructions. Light emission was captured with Amersham Hyperfilm (GE Healthcare) autoradiography film and developed using a Xograft compact X4 developer. Rabbit anti-human PrimPol primary antibody was previously purified in the Doherty lab and used at a dilution of 1/1000. Anti-Rabbit secondary antibody conjugated with horseradish peroxidase (HRP) (Abcam) was used at a dilution of 1/3000. Anti-6xHis Tag conjugatd with HRP (Abcam) primary antibody was used at a dilution of 1/5000.

2.3.2. Annealing of Primer-Template Substrates

The DNA oligonucleotides used in primer extension assays, polymerase processivity assays, electrophoretic mobility shift assays (EMSAs), and single turnover kinetic assays are detailed in Table 2.6. Prof. Shigenori Iwai donated DNA containing a (6-4)PP. ATDBio or Sigma-Aldrich manufactured all other DNA oligonucleotides.

Oligonucleotides were annealed to a final concentration of 200 nM in annealing buffer (50 mM NaCl, 1 mM EDTA, 20 mM Tris-HCl (pH 7.5)) to a final volume of 500 µL. Oligonucleotides were heated to 95°C for 10 minutes and allowed to return to room temperature to allow annealing. Annealed oligonucleotides were stored at -20°C.

#	Primer	Direction	Label	Sequence
1	Undamaged substrate (50/20mer)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTC ACAGCAGACAAGCCAGCAAGCCAGAAGTTCCGACAACACGCGGGACGCGC
2	50mer ssDNA	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTCGGTCTTCAAGGCTGTTGTGCGCCCTGCGCG
3	50mer dsDNA	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCAAGGCTGTTGTGCGCCCTGCGCG ACAGCAGACAAGCCAGCAAGCCAGAAGTTCCGACAACACGCGGGACGCGC
4	Poly(dA)	5’-3’	5’-Biotin	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
5	Poly(dC)	5’-3’	5’-Biotin	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
6	Poly(dG)	5’-3’	5’-Biotin	GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
7	Poly(dT)	5’-3’	5’-Biotin	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
8	Single-incorporation kinetics substrate	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCA ACAGCAGACAAGCCAGCAAGCCAGAAGTTCCGACAACACGCGGGACGCGC
9	Fidelity substrate (AA)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTC ACAGCAGACAAGCCAGCAAGCCAGAAGAACCGACAACACGCGGGACGCGC
10	Fidelity substrate (CC)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTC ACAGCAGACAAGCCAGCAAGCCAGAAGCCTTGACAACACGCGGGACGCGC
11	Fidelity substrate (GG)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTC ACAGCAGACAAGCCAGCAAGCCAGAAGGGCCGACAACACGCGGGACGCGC
12	Fidelity substrate (TT)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTC ACAGCAGACAAGCCAGCAAGCCAGAAGTTCCGACAACACGCGGGACGCGC
13	Mismatch substrate (A/AA)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCA ACAGCAGACAAGCCAGCAAGCCAGAAGAACCGACAACACGCGGGACGCGC
14	Mismatch substrate (C/AA)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCC ACAGCAGACAAGCCAGCAAGCCAGAAGAACCGACAACACGCGGGACGCGC
15	Mismatch substrate (G/AA)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCG ACAGCAGACAAGCCAGCAAGCCAGAAGAACCGACAACACGCGGGACGCGC
16	Mismatch substrate (T/AA)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCT ACAGCAGACAAGCCAGCAAGCCAGAAGAACCGACAACACGCGGGACGCGC
17	Mismatch substrate (A/CC)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCA ACAGCAGACAAGCCAGCAAGCCAGAAGCCCCGACAACACGCGGGACGCGC
18	Mismatch substrate (C/CC)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCC ACAGCAGACAAGCCAGCAAGCCAGAAGCCCCGACAACACGCGGGACGCGC
19	Mismatch substrate (G/CC)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCG ACAGCAGACAAGCCAGCAAGCCAGAAGCCCCGACAACACGCGGGACGCGC

20	Mismatch substrate (T/CC)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCT ACAGCAGACAAGCCAGCAAGCCAGAAGCCCCGACAACACGCGGGACGCGC
21	Mismatch substrate (A/GG)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCA ACAGCAGACAAGCCAGCAAGCCAGAAGGGCCGACAACACGCGGGACGCGC
22	Mismatch substrate (C/GG)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCC ACAGCAGACAAGCCAGCAAGCCAGAAGGGCCGACAACACGCGGGACGCGC
23	Mismatch substrate (G/GG)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCG ACAGCAGACAAGCCAGCAAGCCAGAAGGGCCGACAACACGCGGGACGCGC
24	Mismatch substrate (T/GG)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCT ACAGCAGACAAGCCAGCAAGCCAGAAGGGCCGACAACACGCGGGACGCGC
25	Mismatch substrate (A/TT)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCA ACAGCAGACAAGCCAGCAAGCCAGAAGTTCCGACAACACGCGGGACGCGC
26	Mismatch substrate (C/TT)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCC ACAGCAGACAAGCCAGCAAGCCAGAAGTTCCGACAACACGCGGGACGCGC
27	Mismatch substrate (G/TT)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCG ACAGCAGACAAGCCAGCAAGCCAGAAGTTCCGACAACACGCGGGACGCGC
28	Mismatch substrate (T/TT)	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCT ACAGCAGACAAGCCAGCAAGCCAGAAGTTCCGACAACACGCGGGACGCGC
29	CPD Substrate 1	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTC ACAGCAGACAAGCCAGCAAGCCAGAAGT=TCCGACAACACGCGGGACGCGC
30	CPD Substrate 2	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTC ACAGCAGACAAGCCAGCAAGCCAGAAGT=TCCGACAACACGCGGGACGCGC
31	CPD Substrate 3	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTCGGTCTTCA ACAGCAGACAAGCCAGCAAGCCAGAAGT=TCCGACAACACGCGGGACGCGC
32	(6-4)PP Substrate 1	5’-3’ 3’-5’	5’-Hex None	CACTGACTGTATGATG GTGACTGACATACTACT^TCTACGACTGCTC
33	(6-4)PP Substrate 2	5’-3’ 3’-5’	5’-Hex None	CACTGACTGTATGATGT GTGACTGACATACTACT^TCTACGACTGCTC
34	(6-4)PP Substrate 3	5’-3’ 3’-5’	5’-Hex None	CACTGACTGTATGATGA GTGACTGACATACTACT^TCTACGACTGCTC
35	(6-4)PP Substrate 4	5’-3’ 3’-5’	5’-Hex None	CACTGACTGTATGATGC GTGACTGACATACTACT^TCTACGACTGCTC
36	8-oxo-G Substrate 1	5’-3’ 3’-5’	5’-Hex None	TGTCGTCTGTTCGGTCGTTC ACAGCAGACAAGCCAGCAAGCCAGAAG8oGCCGACAACACGCGGGACGCGC

37	8-oxo-G Substrate 2	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTCGGTCTTC
		3’-5’	None	ACAGCAGACAAGCCAGCAAGCCAGAAG8oGCCGACAACACGCGGGACGCGC
38	Tg Substrate 1	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTC
		3’-5’	None	ACAGCAGACAAGCCAGCAAGCCAGAAGTgCCGACAACACGCGGGACGCGC
39	Uracil Substrate 1	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTC
		3’-5’	None	ACAGCAGACAAGCCAGCAAGCCAGAAGUCCGACAACACGCGGGACGCGC
40	Uracil Substrate 2	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTCGGTCTTC
		3’-5’	None	ACAGCAGACAAGCCAGCAAGCCAGAAGUCCGACAACACGCGGGACGCGC
41	AP Substrate 1	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTC
		3’-5’	None	ACAGCAGACAAGCCAGCAAGCCAGAAGAPCCGACAACACGCGGGACGCGC
42	Fidelity Substrate (CCCC)	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTCGGTCTTC
		3’-5’	None	ACAGCAGACAAGCCAGCAAGCCAGAAGCCCCGACAACACGCGGGACGCGC
43	AP Substrate 2	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTCGGTCTTC
		3’-5’	None	ACAGCAGACAAGCCAGCAAGCCAGAAGAPCCGACAACACGCGGGACGCGC
44	Tg Substrate 2	5’-3’	5’-Hex	TGTCGTCTGTTCGGTCGTTCGGTCTTC
		3’-5’	None	ACAGCAGACAAGCCAGCAAGCCAGAAGTgCCGACAACACGCGGGACGCGC
Crystallography Substrates				
C1	Crystallography Substrate 1	5’-3’	None	GCGTCAT
		3’-5’	None	CGCAGTATTACT

Table 2.6. List of oligonucleotide substrates.
The oligonucleotide substrates used in the experimental analysis of PrimPol. Mismatched bases are coloured in orange (i.e. those that do not conform to Watson-Crick base pairs) and damaged bases are coloured in red.

2.3.3. Primer Extension Assays

Primer extension assays were based on a previously described method (Jozwiakowski and Connolly, 2011). Hex-labelled DNA primers were annealed to oligomer templates (see Table 2.6). Reactions were incubated at 37°C with 20 nM DNA, 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$, 1.0mM DTT and 250 µM dNTPs to a final volume of 20 µL unless otherwise specified. Protein concentrations in the reactions were 100 nM for wild-type and C-terminal truncated constructs of PrimPol, and 34 nM for the polymerase domain of PrimPol alone. Control lanes contained no enzyme. Unless otherwise specified, time points were taken at 0.5, 1, 3, 5 and 60 minutes. The reactions were terminated by the addition of 2x stop buffer (95% formamide, 0.09% xylene cyanol, 0.05% bromophenol blue, 200 nM competitor oligonucleotide) and boiled at 95°C for 5 minutes. Samples were resolved by electrophoresis on a 15% polyacrylamide gel (37.5% (v/v) acrylamide/bisacrylamide (19:1) 40% mix (National Diagnostics), 7 M urea, TBE (89 mM Tris, 2mM EDTA, 0.89 M boric acid, pH 8.3) 0.1% (w/v) APS, 0.03% (v/v) TEMED). Gels were pre-run at a constant power of 15 W (~850 V) for 30-60 minutes. Samples were loaded and electrophoresed at 850V for 2.5 hours in TBE buffer. Fluorescently labelled DNA oligomers were detected by scanning using a Fujifilm FLA-5100 image reader.

2.3.4. Polymerase Processivity Assays

Polymerase processivity assays were based on a previously described method (Washington et al., 1999). Hex-labelled DNA primers were annealed to oligomer templates (see Table 2.6). Protein was preincubated at 37°C with 100 nM protein, 60nM DNA, 10 mM Bis-Tris-Propane-HCl (pH 7.0), 1 0mM $MgCl_2$ and 1.0 mM DTT for one hour. Reactions were initiated through the addition of 250 µM dNTPs and an excess of sonicated herring sperm DNA (1 mg mL^{-1}) as an enzyme trap. Reactions were quenched at time points of 15, 30, 60, 120 and 360 seconds through the addition of 2x stop buffer. To test the effectiveness of the trap, each of the enzymes were also preincubated with excess of herring sperm DNA (1 mg mL^{-1}), as well as 60 nM DNA, 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$ and 1 mM DTT and reactions initiated with the addition of 250 µM dNTPs. To ensure the trap did not disrupt the processivity of the enzyme, the experiment was also carried out with the Klenow fragment of Taq polymerase. Samples were resolved by electrophoresis on a 15% polyacrylamide gel as with primer extension assays and fluorescently labelled DNA

oligomers were detected by scanning using a Fujifilm FLA-5100 image reader. The band intensities were measured using ImageQuant software (GE Healthcare). The percentage of active polymerase at a given nucleotide position is given by the following equation:

$$\% \text{ active polymerases at } n = (I_n + I_{n+1} + I_{n+2}...) \times 100\% / (I_1 + I_2 + I_3...)$$

Where I_1 is the intensity at position 1, I_n is the intensity at position *n* and so on.

2.3.5. Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays (EMSAs) were carried out on hex-labelled DNA primers, either single-stranded or annealed to oligomer templates (see Table 2.6). Varying concentrations of proteins were added to 40-60 nM DNA, 10mM Bis-Tris-Propane-HCl (pH 7.0), 10mM $MgCl_2$ and 1.0 mM DTT to a final volume of 20 µL and incubated at 25°C for 60 minutes. The reactions were supplemented with 2 µL 25% (w/v) ficoll. Samples were resolved by electrophoresis on a 5% polyacrylamide gel (12.5% (v/v) acrylamide/bisacrylamide (19:1) 40% mix (National Diagnostics), 0.5x TBE (44.5 mM Tris, 1mM EDTA, 0.445 M boric acid, pH 8.3), 0.1% (w/v) APS, 0.03% (v/v) TEMED) cast in a 1 mm Novex Gel cassette (Invitrogen). Gels were run at 75 V for 1 hour in 0.5x TBE buffer in the XCell SureLock Mini-Cell Electrophoresis System (Invitrogen). Fluorescently labelled DNA oligomers were detected by scanning using a Fujifilm FLA-5100 image reader.

2.3.6. Single Turnover Kinetic Assays

The single turnover assays were based on a previously described method (Nemec et al., 2012). Hex-labelled fluorescent DNA (Table 2.6) was incubated with 200 nM protein. The DNA/protein mixture was incubated with varying concentrations of dATP for time points ranging from 0.5-60 minutes and each reaction was carried out in triplicate. Reactions were quenched by the addition of 2x stop buffer. The products were resolved by electrophoresis as in the primer extension assays (Section 2.3.3). The concentrations of extended and unextended fluorescent DNA primers were measured using ImageQuant software (GE Life Sciences). The concentration of extended product relative to total fluorescent DNA loaded was plotted as a function of time and the data were fit to a single-exponential curve to obtain the k_{obs} for the different dATP concentrations using the following exponential equation:

$$[product] = A(1 - e^{-k_{obs}t})$$

A secondary plot of k_{obs} as a function of dATP concentration was fitted to the following hyperbolic equation:

$$k_{obs} = \frac{k_{pol}[dATP]}{K_{d(dNTP)} + [dATP]}$$

where k_{pol} is the rate of polymerisation and $K_{d(dNTP)}$ is the equilibrium dissociation constant for dATP.

2.3.7. Analysis of Protein Interaction by Gel Filtration

Protein interaction was analysed by size-exclusion chromatography on a Superdex S75 10/30 GL gel filtration column (GE Healthcare). The column was calibrated using albumin (67,000 Da), ovalbumin (43,000 Da), chymotrypsinogen A (25,000 Da), ribonuclease A (13,700 Da), and aprotinin (6,512). The protein was loaded at 0.5 mL min^{-1}. Retention volume of the protein was plotted against the molecular weight of the protein to reliably predict protein molecular weights. The column was pre-equilibrated in a buffer containing 50 mM Tris-HCl, 100 mM NaCl and 2mM TECEP that had been sterile-filtered using a 0.2 µm pore size vacuum filtration system (Nalgene). 0.5 mL of protein was loaded at a concentration of 35 µM. Protein-containing fractions were determined by A_{280} level and samples of these fractions were subjected to SDS-PAGE analysis (see Section 2.3.1).

2.4. Biophysical Methods

2.4.1. Inductively Coupled Plasma Mass Spectrometry

The concentrations of zinc ions in protein samples were determined using inductively coupled plasma mass spectrometry (ICP-MS). For this analysis of human PrimPol, the proteins were gel filtered in zinc-free solutions to minimise the concentration of zinc ions that were not chelated by protein. Quantitative calibrations were made using standards prepared from zinc solutions containing 0, 100, 500 and 1000 ng mL^{-1}. Each sample was measured in triplicate and background measurements of buffer without protein were determined and subtracted from the appropriate readings. A wash was performed between standards and samples using 2% nitric acid.

2.4.2. In-gel Digestion and Mass Spectrometry

Samples for mass spectrometry analysis were resolved by SDS-PAGE (Section 2.3.1) and Coomassie stained (Section 2.3.1.1). Bands of interest were excised and destained with a solution of 50% acetonitrile and 25 mM ammonium bicarbonate whilst vortexing. Bands were dehydrated for 5 minutes using a speedvac. Bands were rehydrated with 50µL of 10 mM DTT in 25 mM ammonium bicarbonate for 45 minutes at 50°C to reduce proteins. Bands were then allowed to return to room temperature and excess DTT solution was removed and replaced with 500 mM iodoacetamide in 25 mM ammonium bicarbonate and this was incubated for 45 minutes in the dark. Bands were subsequently dehydrated as before and rehydrated in a 1 µg μL^{-1} trypsin solution and left on ice for 5 minutes. Excess trypsin was removed and bands were covered with 25 mM ammonium bicarbonate and incubated overnight at 37°C. Formic acid was then added to the gel bands to a final concentration of 5% (v/v). Bands were vortexed, centrifuged on a benchtop centrifuge at 13,000 rpm for 10 minutes and the supernatant was extracted to a separate tube. This was repeated with one volume of acetonitrile. The volume of protein solution was reduced to ~8 µL through concentration by speedvac. Protein solutions were stored at -20°C until required for mass spectrometry.

Samples were analysed by mass spectrometry on an LTQ Orbitrap XL. Mass spectrometry was scanned across a mass/charge ratio of 200-2,000 and data was acquired using Xcalibur 2.1 software (Thermo Scientific). Proteins were loaded at a flow rate of 0.3 µL min^{-1} in a buffer of 0.1% formic acid, 1% acetonitrile onto a reverse phase HPLC column and eluted using a gradient up to 20% acetonitrile in 0.1% formic acid over a period of 40 minutes. Eluted peptides were analysed using a spray voltage of 2.2-3.0 kV. 2 µL of initial BSA standard (25 fmol μL^{-1}) was loaded prior to individual protein samples. Raw files were converted into mgf format using proteomic analysis software Compass (Wenger et al., 2011) and proteins were identified from primary sequence databases using Mascot (Perkins et al., 1999).

2.4.3. Circular Dichroism Spectroscopy

Samples were buffer exchanged into a buffer of 400 nM NaF, 15 mM Tris-HCl (pH 7.5) and concentrated to 3 mg mL^{-1}. The concentration of recombinant protein was determined before and after the circular dichroism (CD) scans, to ensure accurate secondary structure predictions. Samples were placed in a 0.1 mm quartz cuvette (Starna, Essex, UK) and measurements were taken using a JASCO J-715

spectropolarimeter (JASCO). The CD spectrum of the empty buffer was subtracted from that of the sample and time constant was set to 4 s with a scan rate of 50 nm min^{-1}. The bandwidth was 1 nm and the sensitivity set to standard. Scans were performed in triplicate from 320 to 180 nm with a 0.1 nm data pitch and continuous scan mode. A Peltier device was used to maintain a temperature at 20°C.

2.4.4. Thermal Denaturation

Proteins were buffer exchanged into a buffer of 200 mM NaF, 15 mM Tris-HCl (pH7.5). 45 µl of 1 µM protein solution was added to 15 µl of SYPRO Orange, resulting in a final protein concentration of 0.75 µM. A control of 45 µl buffer was also added to 15 µl of SYPRO Orange. 20 µl of each sample was aliquoted in triplicate into an opaque 96 well plate. Protein melting experiments were carried out using the LightCycler 480 System II (Roche). The instrument was configured with a detection format of 465 nm as the wavelength of excitation and 580 nm as the emission wavelength to detect SYPRO Orange-specific signal. Denaturation curve fluorescent signal was acquired within a range of 20°C to 80°C using a ramping rate of 0.03°C s^{-1} and an acquisition of 20 data points per degree Celsius. Melting temperatures (T_m) were determined through the measurement of the lowest point of the negative differential of the denaturation curve. Data was corrected for the background signal of the buffer conditions and presented as units of fluorescence with respect to temperature ± 1 standard deviation.

2.4.5. Protein Crystallisation

2.4.5.1. Pre-crystallisation Trials

Proteins were buffer exchanged into a suitable buffer for crystallography, typically to remove glycerol or to lower salt concentration. Pre-crystallisation trials were performed for each protein prior to crystallographic screening to estimate a suitable protein concentration for screening. Trials were constructed using the Pre-Crystallization Test (PCT) (Hampton Research) according to the manufacturer's instructions.

2.4.5.2. Crystallographic Screening

Crystallographic screening was conducted using a Phoenix liquid handling system (Art Robbins Instruments) and a number of optimised crystallisation screens. These screens include Crystal Screen I (Hampton Research), Structure Screen II (Molecular Dimensions), PEGs Suite (Qiagen), MPD Suite (Qiagen), Ammonium Sulphate Screen (Qiagen), Nucleix Suite (Qiagen), Stura Footprint Screen (Molecular Dimensions),

MacroSol (Molecular Dimensions), ProPlex (Molecular Dimensions), and MIDAS (Molecular Dimensions). Crystallisation screens were set up in a sitting drop format with a 0.35 µL protein solution typically mixed 1:1 with 0.35 µL reservoir condition, with reservoir volume of 75 µL. Screens were sealed and incubated at a fixed temperature, generally 12°C.

Crystallography screening by capillary diffusion was also set up using the Crystal HARP screen (Molecular Dimensions) according to the manufacturer's instructions. This was incubated at 12°C.

2.4.6. Determination of Protein Atomic Structure by X-ray Diffraction

2.4.6.1. Protein Crystal Cryogenic Protection

Crystal samples were mixed with cryo-protectant to protect crystals from ice crystals upon cooling. Protein crystals were looped from the drop using 100-200 µm aperture Micro Mounts (MeTeGen) and wafted into a drop of mother liquor containing cryo-protectant, typically 2-methyl-2,4-pentanediol (MPD) or ethylene glycol.

2.4.6.2. Crystallographic Analysis

Following cryo-protection, crystals were mounted onto the goniometer of an X-ray generator. Single wavelength diffraction data were collected for a number of crystals using Rigaku MicroMax 007-HF (University of Sussex), and an RPA70N-PrimPol$_{480\text{-}560}$ crystal at station I03 Diamond Light Source, Didcot, UK. Diffraction data was analysed for protein content at perpendicular ϕ angles of 0° and 90°. Given sufficient quality data, a larger range data set was collected for crystals. The diffraction data were processed with iMOSFLM (Battye et al., 2011) and programs from the CCP4 suite (Collaborative Computational Project, Number 4, 1994 (E. Potterton et al., 2002; L. Potterton et al., 2004)).

Molecular replacement and refinement was carried out by Dr. Nigel Brissett. All structural images were prepared using CCP4mg (McNicholas et al., 2011). A LigPlot schematic of the interactions between PrimPol and RPA70N was produced using the LigPlot^{+} software (Laskowski and Swindells, 2011).

2.5. Bioninformatic and Analytic Tools

2.5.1. Secondary Structure Prediction Tools

Secondary structure predictions of PrimPol were made using the University College London PSIPRED 2.0 server (Buchan et al., 2013; Jones, 1999). The secondary structure prediction of human PrimPol is shown in Appendix I.

2.5.2. Identification of Protein Homologues

Protein homologues were identified using the NCBI position-specific iterative basic local alignment search tool (PSI-BLAST) (Altschul et al., 1997).

2.5.3. Gene Sequence Acquisition and Alignment

Gene sequences were retrieved using the National Centre for Biotechnology Information from GenBank (Benson et al., 2009; Sayers et al., 2010). Multiple sequence alignments were created using Clustal Omega (Sievers et al., 2011) and analysed using JalView and CLC Sequence Viewer (CLCBio, 2014; Waterhouse et al., 2009).

2.5.4. Computation of Physical and Chemical Parameters of Proteins

The program ProtParam, from the ExPASy server, was used for gathering the physical and chemical parameters of proteins as estimated from their primary protein sequence including pI, molecular weight and molar extinction coefficient (Gasteiger et al., 2005).

2.5.5. Analysis of DNA fibre length from DNA combing experiments

DNA fibre progression and immunolabelling experiments were carried out as described previously (Edmunds et al., 2008) by Dr. Laura Bailey. The data generated from these experiments were analysed using the OMERO 5 software package (The Open Microscopy Environment). Jackson and Pombo calculate that 1 µm represents 2.59 kb of DNA (Jackson and Pombo, 1998) and given a fixed length of time, in our case 20 minutes, the rate of synthesis can be calculated from the length of the fibres in µm as follows:

$$Replication\ Rate\ (kb\ min^{-1}) = \frac{Length\ of\ fibre(\mu m) \times Base\ pairs\ per\ \mu m\ (2.59\ kb\ \mu m^{-1})}{Time\ (min)}$$

CHAPTER 3

Dissecting the Domain Architecture of Eukaryotic PrimPol

3.1. Introduction

Until recently, the Pol α-associated DNA primase small subunit (PriS) responsible for *de novo* polymer synthesis through the production of RNA primers was considered to be the sole AEP superfamily member present in eukaryotes (Frick and Richardson, 2001). However, bioinformatic analysis identified the existence of an additional uncharacterized DNA primase in eukaryotes called PrimPol, which belongs to the 'NCLDV-herpesvirus clade' of viral AEPs (L. M. Iyer et al., 2005). Recent studies have reported that PrimPol is a DNA primase, with the ability to synthesise primers using either rNTPs or dNTPs with a preference for DNA primer production (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Keen et al., 2014; Mourón et al., 2013; Rudd et al., 2013; Wan et al., 2013).

Comparative analysis of the amino acid sequence of eukaryotic PrimPols identified the presence of two distinct domains (L. M. Iyer et al., 2005), an enzymatic AEP polymerase domain and a UL52-like zinc finger domain (see Figure 1.22). Mutation of the conserved zinc-chelating cysteine residues in the UL52 subunit of herpes simplex virus resulted in a severe reduction in DNA binding (Biswas and S. K. Weller, 1999; Y. Chen et al., 2005). The AEP polymerase domain of PrimPol contains the three signature catalytic motifs (see Figure 1.22); a highly conserved DxE motif (motif I) together with an aspartate residue from motif III forms the divalent metal binding site, and the SxH of motif II forms part of the putative nucleotide binding motif. Mutagenesis of these motifs abolishes the catalytic activities of these polymerases (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Lao-Sirieix and S. D. Bell, 2004; Lipps et al., 2004; Rudd et al., 2013).

In this chapter, the molecular architecture of PrimPol is dissected to define the activities associated with its two major functional domains. PrimPol from Homo sapiens (human), X. tropicalis and Mus musculus (mouse) are examined (Figure 3.1). We demonstrate that the zinc finger domain is a single-stranded DNA binding motif that is crucial for primase activity. Polymerase activity, on the other hand, appears to be largely independent of the zinc finger domain. The AEP polymerase domain, responsible for both polymerase and primase activity is determined to bind both single-stranded and double-stranded DNA, as it would at a primer-template junction.

3.2. Expression and Purification of Wild-type PrimPol

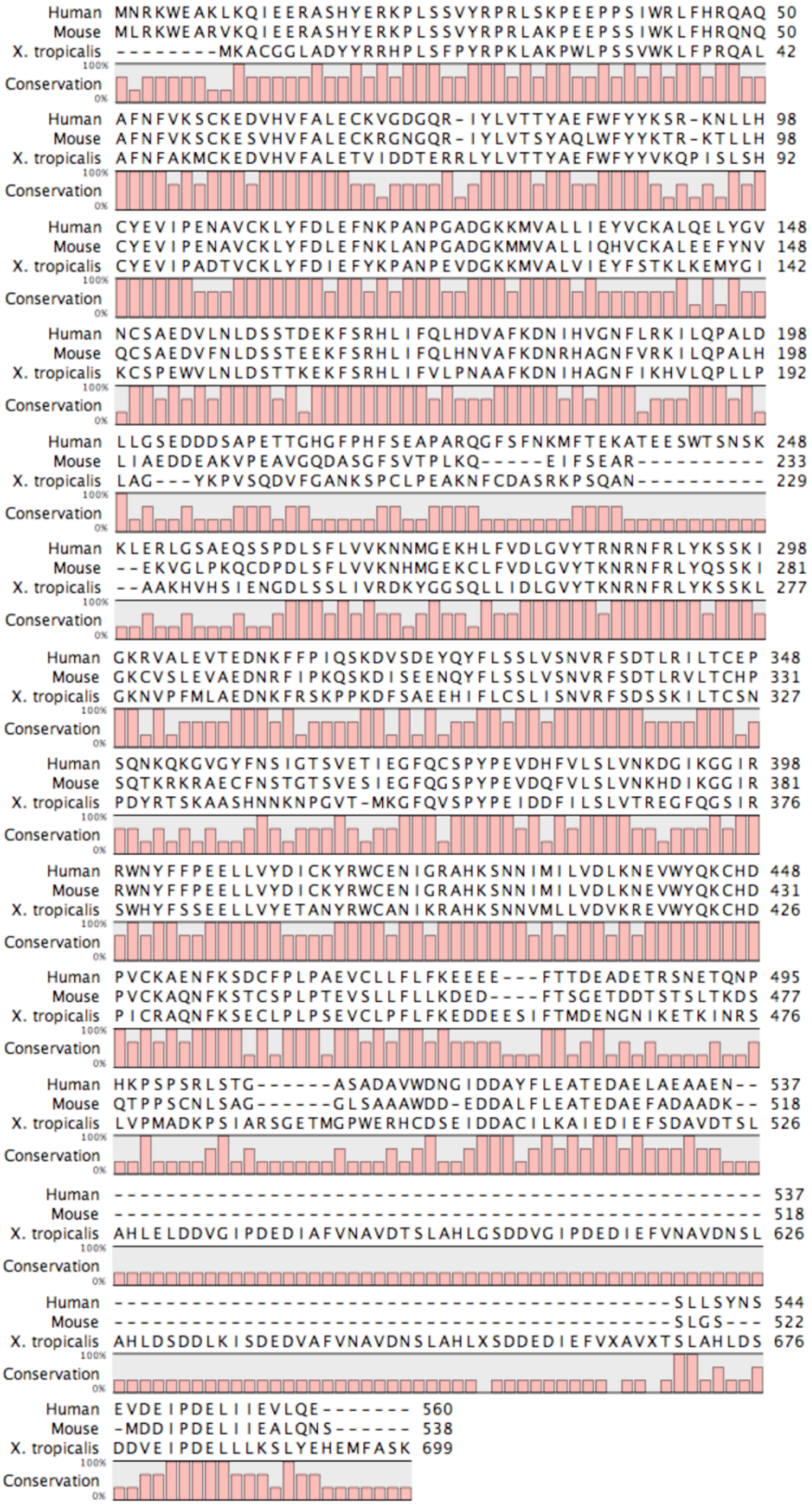


Figure 3.1. Alignment of eukaryotic PrimPol homologues.
The alignment of the three PrimPol homologues that are analysed biochemically in this thesis from *Homo sapiens* (Human), *Mus musculus* (Mouse), and *Xenopus tropicalis*.

The PrimPol constructs that were produced to assess their biochemical activities are shown in Figure 3.2. In order to assess the activities of PrimPol domains relative to the wild-type protein, we purified the wild-type PrimPols from both *Homo sapiens* (human) and *X. tropicalis*.

3.2.1. Human PrimPol Purification

The human PrimPol expression construct (plasmid H1, Table 2.3) was provided by my colleague Dr. Julie Bianchi and was expressed and purified as described in Tables 2.4 and 2.5. Previous studies in the lab have shown that human PrimPol shows greatest soluble expression levels in *Escherichia coli* SHuffle competent cells. These SHuffle cells are a variant of the K12 strain of *E. coli* that have been engineered to express molecular chaperones to assist with the folding of proteins with multiple disulphide bonds, including the disulphide bond isomerase DsbC (Bessette et al., 1999; Levy et al., 2001). The pET28a expression vector was utilised for expression as it allowed a (6x-His) Affinity tag to be fused to the N-terminus of PrimPol thus allowing for purification by affinity of the aromatic hydrocarbon rings to nickel affinity resin. PrimPol was eluted from the nickel affinity chromatography column by imidazole (see Section 2.2.4) and the correct size confirmed by SDS-PAGE (Figure 3.3A). PrimPol was further purified using heparin column chromatography, a pseudo-affinity column that acts as a DNA mimetic for proteins that bind nucleic acids (Figure 3.3A).

In addition to the wild-type human PrimPol, a zinc finger knockout (ZF-KO) variant of PrimPol was also constructed (plasmid H11, Table 2.3). The first invariant cysteine and the invariant histidine of the zinc finger (C418 and H425 in human PrimPol) were mutated to alanines by site directed mutagenesis for the construction of this zinc finger knockout. The $PrimPol_{ZF\text{-}KO}$ recombinant protein was purified by my colleague Dr. Stanislaw Jozwiakwoski for use in experiments described in this thesis.

3.2.2. Purification of *Xenopus tropicalis* PrimPol

A homologue of human PrimPol in *X. tropicalis* was purified to allow confirmation of our results using PrimPol from two different organisms, one mammalian and one amphibian. *X. tropicalis* PrimPol (XPrimPol) was identified by the PSI-BLAST program, an amino acid sequence homology search program. This sequence (accession number NP_001015920) is a truncated form of the protein, providing only the N-terminal 180 amino acids. Reannotation of this gene by introducing frameshift mutations close to the

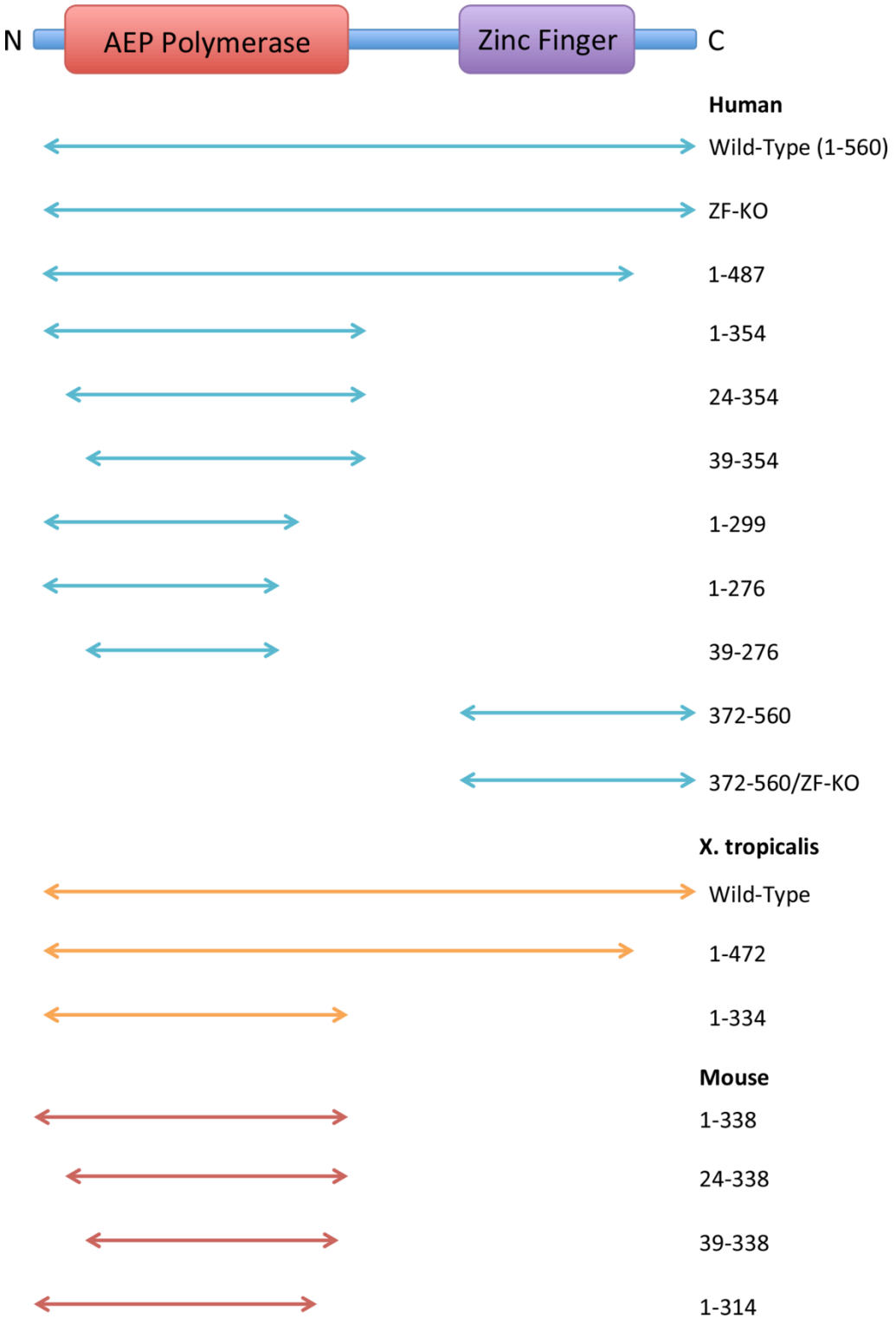


Figure 3.2. Recombinant PrimPol variants.
The human (blue), *X. tropicalis* (orange) and mouse (red) PrimPol variants analysed in this chapter are shown. Truncated PrimPol variants are depicted by their relative length and compared to the domain architecture shown above. Zinc finger knockout variants, denoted by ZF-KO, have two mutations in zinc-coordinating residues, C418A and H425A.

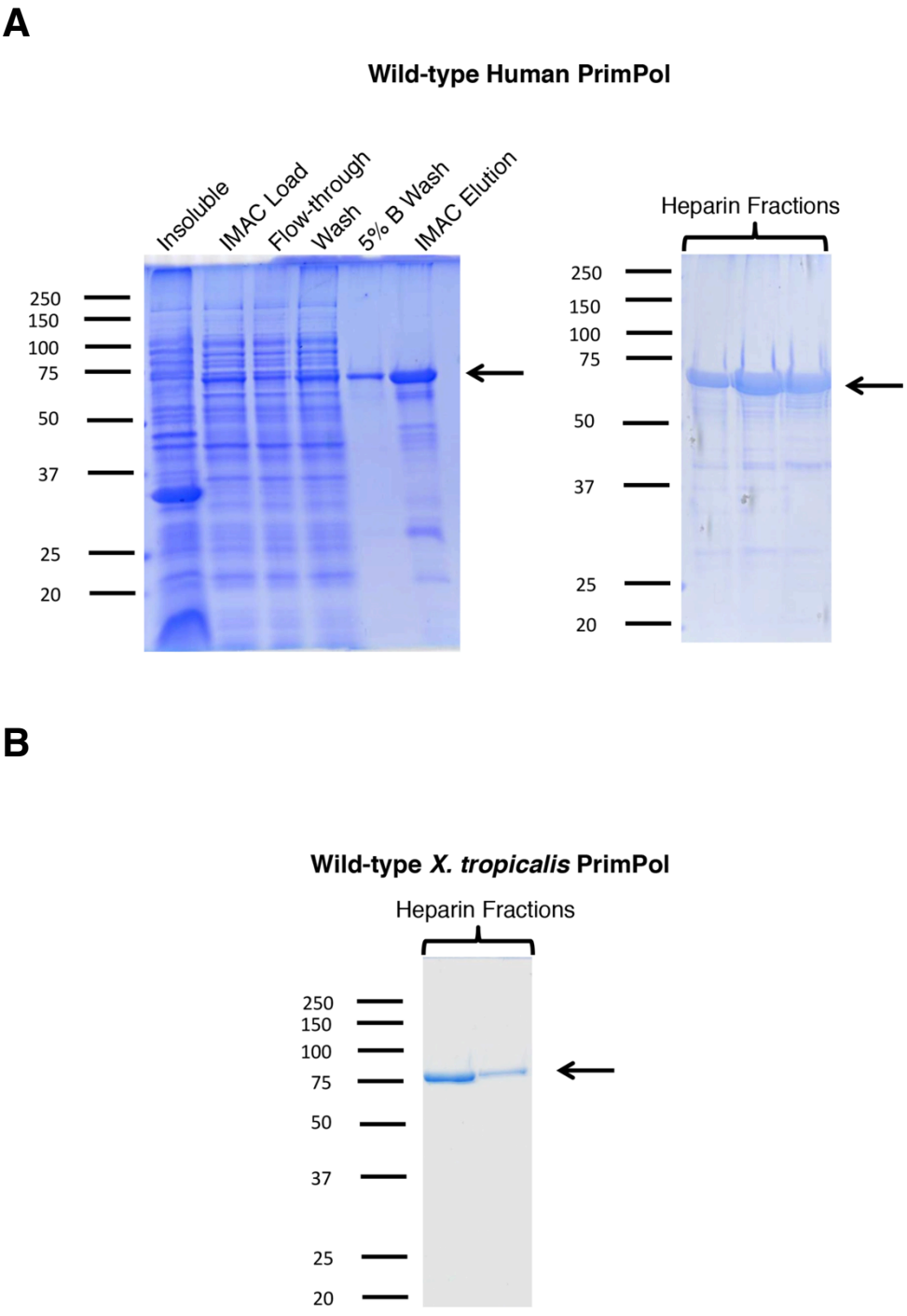


Figure 3.3. The purification of wild-type Human and *Xenopus* PrimPol.
(A) Purification of wild-type Human PrimPol (68.9 kDa). SDS-PAGE analysis of samples taken from the insoluble fraction and soluble fraction (load) from the growth and expression of PrimPol in BL21 as well as IMAC purification samples. The protein-containing fractions from the heparin column are also shown. **(B)** Purified *Xenopus tropicalis* PrimPol (83.5 kDa). SDS-PAGE analysis of the heparin column fractions of *X. tropicalis* PrimPol following IMAC affinity and heparin column purification is shown.

region of the stop codon produced an extended gene product that aligned well with human PrimPol. This reannotated construct (plasmid X1, Table 2.3) was produced through subcloning of a synthetic construct (GenScript). This protein was purified in the same manner as human PrimPol (Figure 3.3B; see also Tables 2.3 and 2.4). This protein is ~83 kDa and this *X. tropicalis* PrimPol was used to produce a polyclonal antibody. This polyclonal antibody was purified and utilised in *in vivo* experiments in *Xenopus laevis* extracts, the protein identified *in vivo* was also an approximate size of ~83 kDa. A more recent alignment of this reannotated *X. tropicalis* PrimPol with an *X. laevis* PrimPol gene (Appendix II) demonstrates that the two proteins align well up the C-terminus of both proteins, validating that our reannotated *X. tropicalis* PrimPol using frameshift mutation is the correct annotation of *X. tropicalis* PrimPol.

3.3. Expression and Purification of PrimPol Truncations

A number of PrimPol truncations were constructed for the biochemical and structural analysis of PrimPol activity in the absence of different parts of the protein. All truncations described in this chapter are depicted in Figure 3.2.

3.3.1. Human PrimPol C-terminal Truncation ($PrimPol_{1-487}$)

The C-terminus of PrimPol, downstream of the zinc finger (amino acids 488-560), was removed by the introduction of a premature stop codon through site-directed mutagenesis (plasmid H8, Table 2.3). This $PrimPol_{1-487}$ construct was also expressed and purified in the same manner as wild-type human PrimPol (Figure 3.4A). The section of the protein C-terminal of the zinc finger domain is not predicted to be well structured and, indeed, removal of this part of the protein increases the protein's stability (see Section 6.2). This $PrimPol_{1-487}$ variant was constructed to determine whether there were any effects on catalytic activity caused by the removal of the C-terminus of PrimPol.

3.3.2. Human PrimPol Polymerase Domain ($PrimPol_{1-354}$)

The predicted polymerase domain of PrimPol (amino acids 1-354) was constructed by PCR from the wild-type variant (plasmid H2, Table 2.3). $PrimPol_{1-354}$ was expressed in BL21 *E. coli* cells, which offered a significantly greater yield of this truncated variant of the PrimPol protein. As this protein was also to be screened for crystallographic studies, additional cation exchange and size-exclusion chromatography steps were

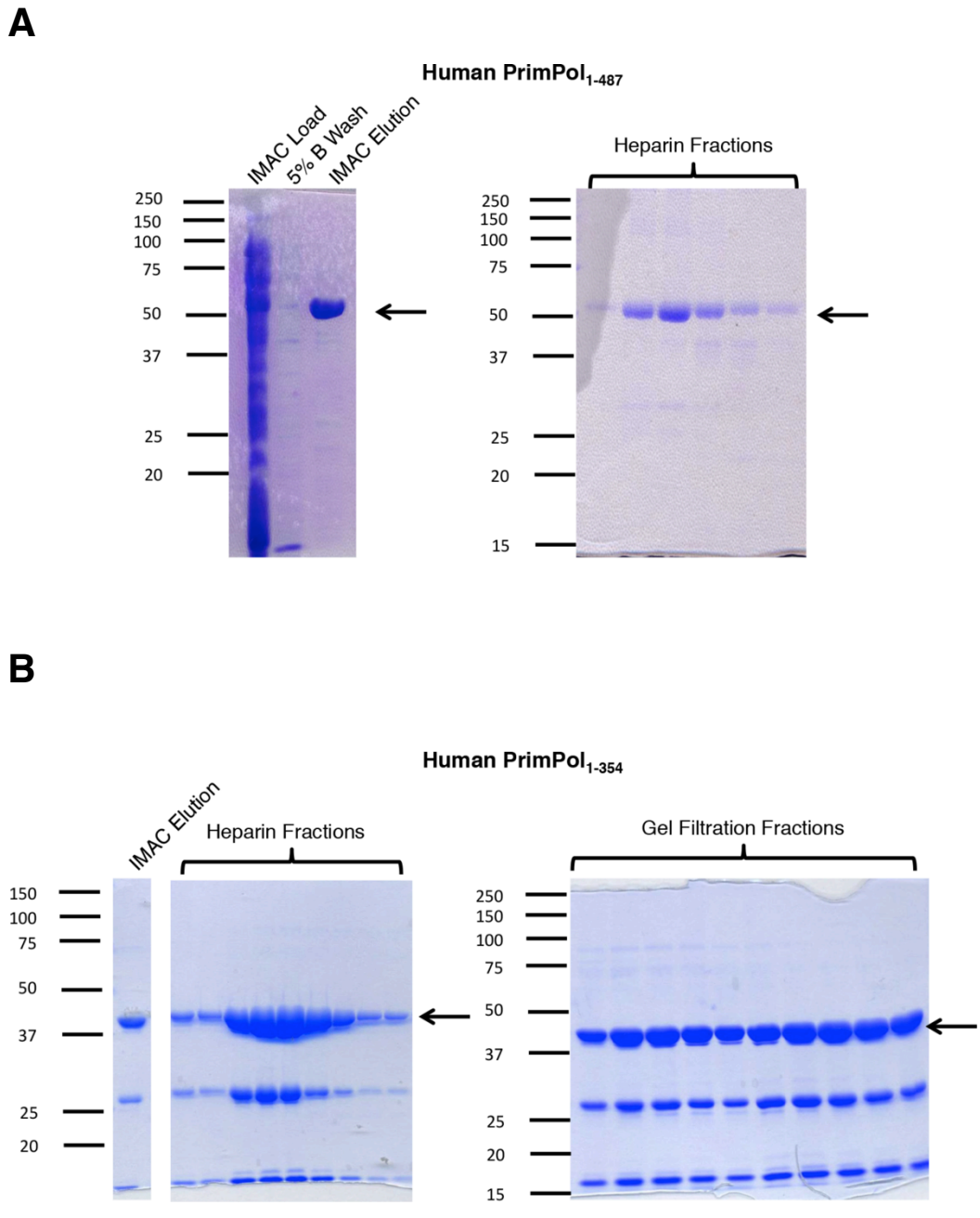


Figure 3.4. Purification of Human PrimPol$_{1-487}$ and PrimPol$_{1-354}$ truncation variants.
(A) SDS-PAGE analysis of the purification of PrimPol$_{1-487}$ (56.4kDa). PrimPol$_{1-487}$ was purified by IMAC affinity and a heparin column. **(B)** SDS-PAGE analysis of the purification of PrimPol$_{1-354}$ (43.2 kDa). PrimPol$_{1-354}$ was purified by IMAC affinity chromatography, heparin column and by size-exclusion chromatography and the protein-containing fractions are shown. The breakdown products evident in the purification of PrimPol$_{1-354}$ are analysed in Section 3.12.1.

carried out and protein purity was determined by SDS-PAGE analysis (Figure 3.4B). As distinctly evident in Figure 3.4B, there is are two proteolysis products at lower molecular weights than the desired product; these well-defined bands were consistently evident and subjected to mass spectrometry to determine their composition (see Section 3.12.1).

3.3.3. Human PrimPol Zinc Finger Domain ($PrimPol_{372-560}$)

The zinc finger domain of PrimPol (amino acids 372-560) was also constructed by PCR from the wild-type variant of PrimPol (plasmid H9, Table 2.3). $PrimPol_{372-560}$ was also expressed in BL21 *E. coli* cells. However as this protein had a predicted pI of 4.32, an anion exchange chromatography column was used to purify this protein following nickel-affinity chromatography (Figure 3.5A).

In addition to the zinc finger domain, a mutated zinc finger knockout variant of the zinc finger domain was constructed ($PrimPol_{372-560/ZF-KO}$). This was constructed by site directed mutagenesis of the 372-560 human PrimPol construct, resulting in the same C418A and H425A mutations as the $PrimPol_{ZF-KO}$ construct. This mutated zinc finger domain was purified in the same manner as the zinc finger domain (Figure 3.5B).

3.3.4. *Xenopus tropicalis* PrimPol Truncations

Two *X. tropicalis* PrimPol truncations were constructed to discern their activities relative to the equivalent human constructs. These *X. tropicalis* PrimPol constructs were amino acids 1-334 and 1-472 (plasmids X2 and X3, Table 2.3). These amino acid positions were determined to be analagous to human 1-354 and 1-487 by sequence alignment (Figure 3.1). The $XPrimPol_{1-472}$ protein was expressed in SHuffle *E. coli* cells and was purified by nickel-affinity chromatography only, but SDS-PAGE analysis illustrates that it was purified to sufficient homogeneity (Figure 3.6A). The $XPrimPol_{1-334}$ protein was expressed in BL21 *E. coli* cells and purified using nickel-affinity chromatography column, a heparin column, and a gel-filtration column (Figure 3.6B).

3.4. PrimPol is an Active Polymerase

Previous studies had shown that PrimPol is indeed an active polymerase (Bianchi et al., 2013; Garcia-Gomez et al., 2013). This was activity was confirmed for both human and *X. tropicalis* PrimPol by standard primer extension assays using unmodified

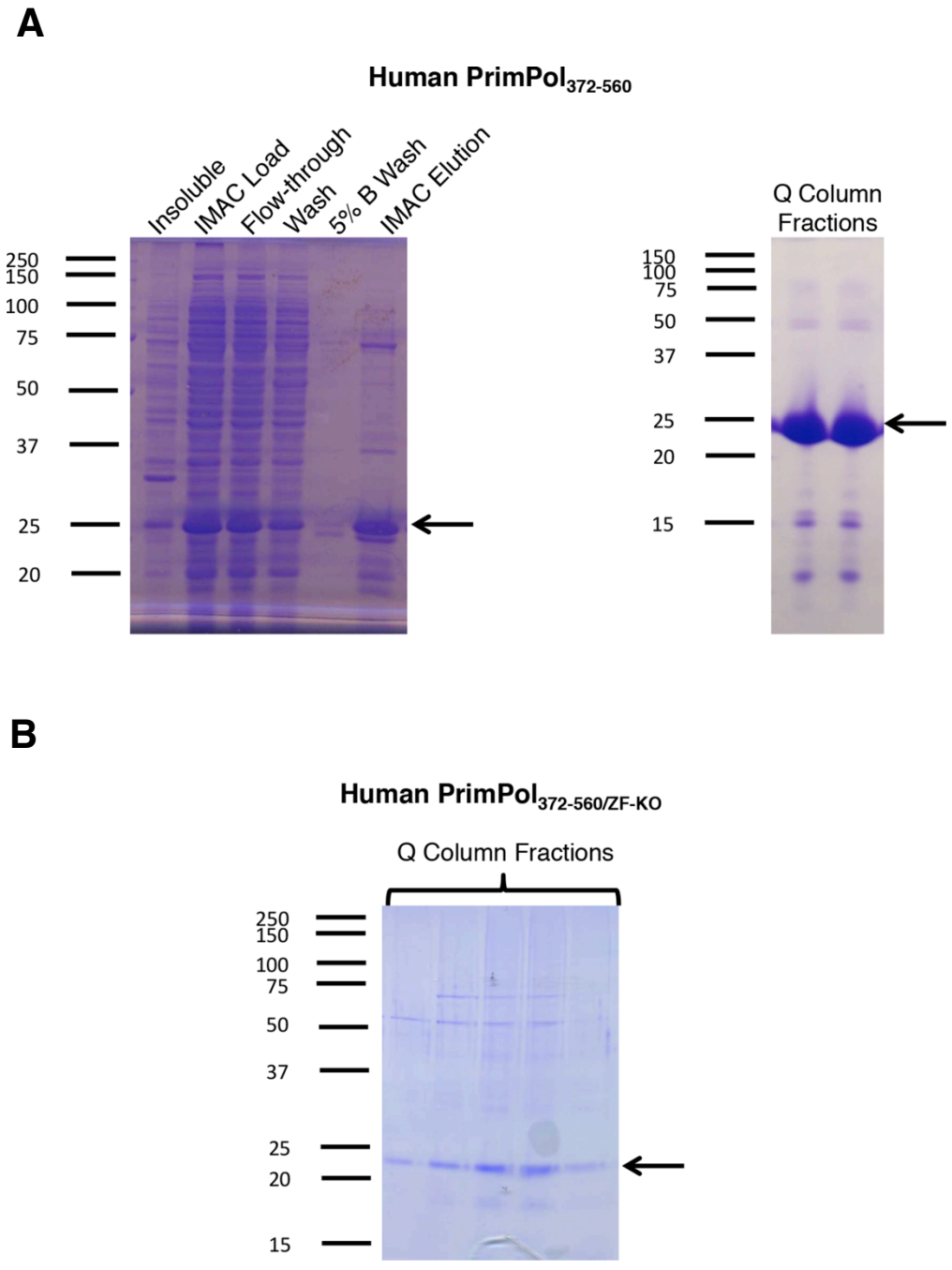


Figure 3.5. Purification of $PrimPol_{372-560}$ and $PrimPol_{372-560/ZF-KO}$.
(A) SDS-PAGE analysis of the purification of $PrimPol_{372-560}$ (24.0kDa). Samples from the insoluble and soluble (load) fractions are analysed along with samples from the IMAC affinity purification of $PrimPol_{372-560}$. Also shown are the samples of the protein-containing fractions from the Q column anion exchange chromatography of $PrimPol_{372-560}$. **(B)** SDS-PAGE analysis of the purification of $PrimPol_{372-560/ZF-KO}$ (23.9kDa). Samples from the Q column anion exchange chromatography of $PrimPol_{372-560/ZF-KO}$ are shown. Not shown here is the IMAC affinity chromatography of $PrimPol_{372-560/ZF-KO}$.

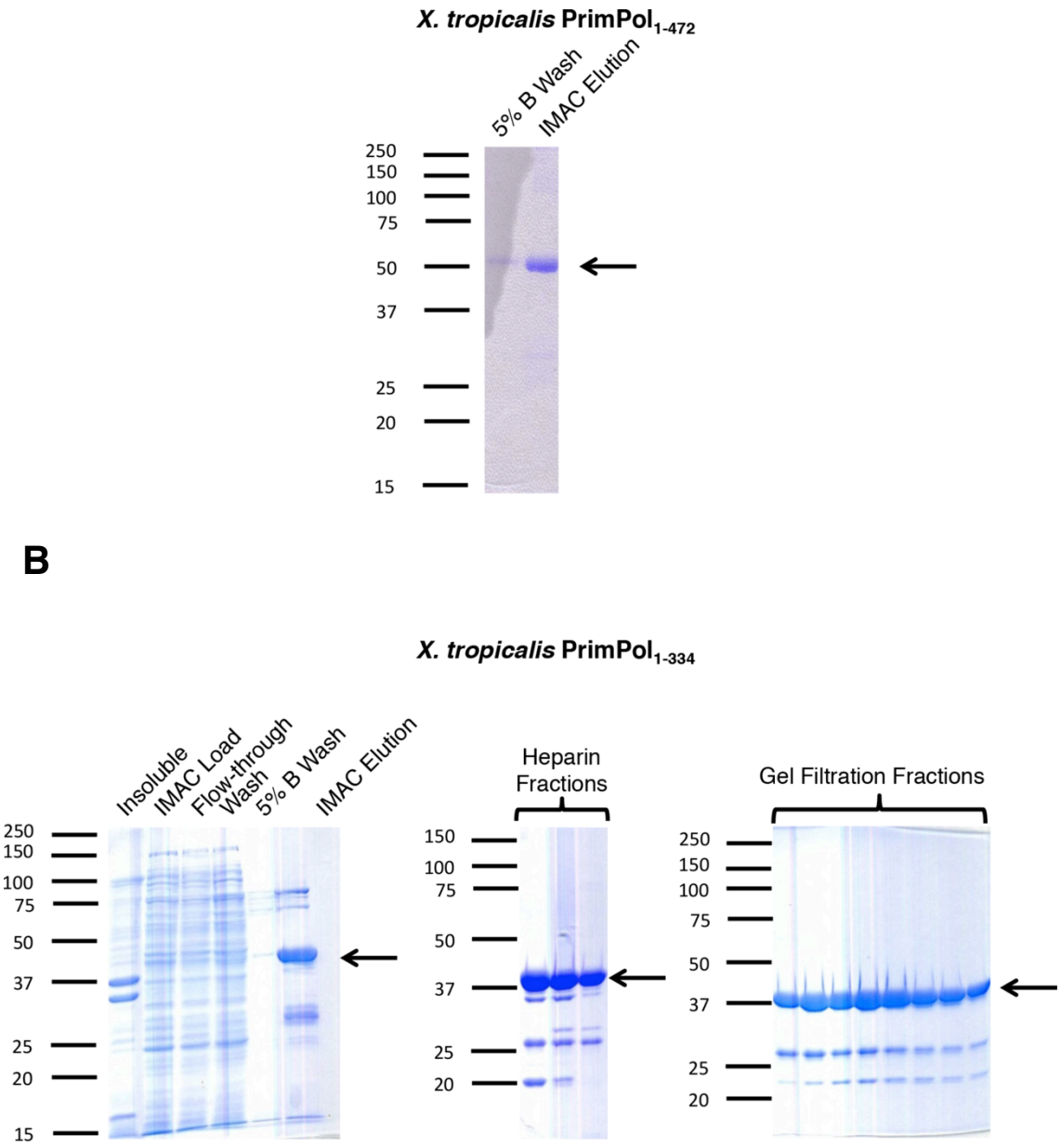


Figure 3.6. Purification of *X. tropicalis* PrimPol truncation variants.
(A) SDS-PAGE analysis of the purification of XPrimPol$_{1-472}$ (56.6 kDa). XPrimPol$_{1-472}$ purified using by IMAC affinity chromatography. **(B)** SDS-PAGE analysis of the purification of XPrimPol$_{1-334}$ (40.5 kDa). XPrimPol$_{1-334}$ was purified through the use of IMAC affinity, heparin column and size exclusion chromatography. The breakdown products evident here are similar to those seen in Figure 3.4B in the purification of Human PrimPol$_{1-354}$ that are analysed in Section 3.12.1.

synthetic primers and dNTPs (Figure 3.7). The polymerase activity of the Klenow fragment (exonuclease deficient) of the polymerase from *Thermus aquaticus* (Klenow Taq) was also tested as a positive control enzyme (Figure 3.7). PrimPol is an active polymerase from both human and *X. tropicalis*, however from this experiment alone it is immediately evident that the polymerase activity of PrimPol is much slower than that of Klenow Taq. Whilst it takes an hour for the PrimPol to fully extend the primer DNA, it takes Klenow Taq less than 30 seconds.

The removal of the unstructured C-terminus of PrimPol gave no reduction in the polymerase activity in both human and *X. tropicalis* systems (Figure 3.8). In fact, in both cases, there appears to be an apparent increase in the activity of both $PrimPol_{1-487}$ and $XPrimPol_{1-472}$ relative to their wild-type counterparts. Fully extended product is evident as early as 3 minutes incubation when the C-terminus is truncated.

3.5. PrimPol Polymerase Activity is Independent of the Zinc Finger Domain

As PrimPol exhibits polymerase activity and it was previously demonstrated that this polymerase activity localises to the AEP polymerase domain of PrimPol (Bianchi et al., 2013), we wanted to ascertain whether the zinc finger domain was simply dispensible for polymerase activity or if it played any crucial ancillary roles. The polymerase activity of PrimPol without the contribution of its putative zinc finger was measured in two ways: the first was a 1-354 truncation mutant that lacked the domain completely and the second was a double point mutant in which zinc-chelating C418 and H425 residues of human PrimPol were mutated to alanine (ZF-KO). Again, a *X. tropicalis* PrimPol truncation equivalent (1-334) was constructed to measure activity from two different organisms' PrimPols.

The removal of the zinc finger did not disrupt the polymerase activity of PrimPol and actually increased its apparent rate of incorporation relative to both wild-type and the C-terminus truncation of PrimPol (Figure 3.9). The concentration of the polymerase domain of PrimPol in these primer extension reactions was 3-fold lower than with the wild-type and C-terminus truncation (34 nM in comparison to 100 nM), yet the dNTP incorporation was comparable or higher in these assays. The $PrimPol_{ZF-KO}$ construct of PrimPol was also an active polymerase but its activity was not increased relative to the wild-type PrimPol (Figure 3.9).

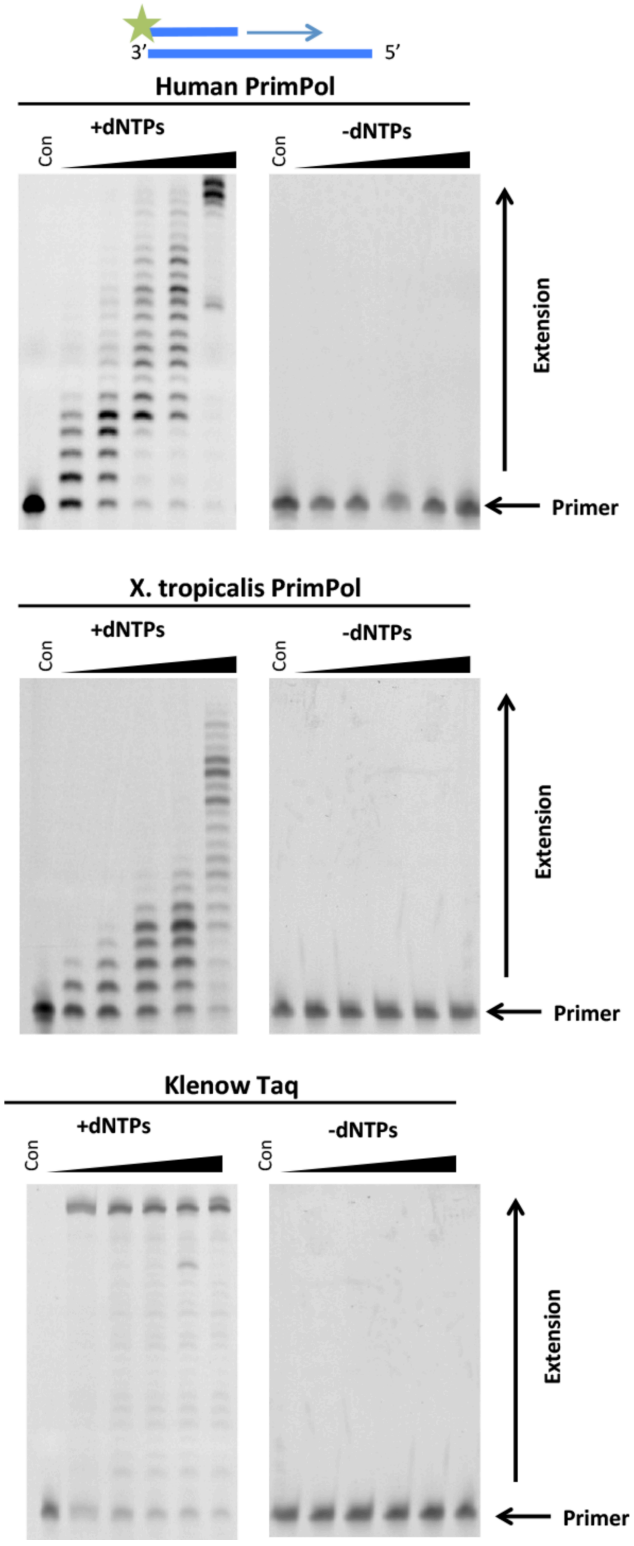


Figure 3.7. Polymerase activity of PrimPol.
Human and *X. tropicalis* PrimPol was incubated with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points, along with a control containing no protein. PrimPol is proficient at extending an undamaged oligonucleotide substrate using dNTPs. The proteins were also analysed in the absence of dNTPs and no activity was observed. As a positive control for polymerase activity, the Klenow fragment of the replicative polymerase from Thermus aquaticus (Klenow Taq), was analysed and significant polymerase activity is evident.

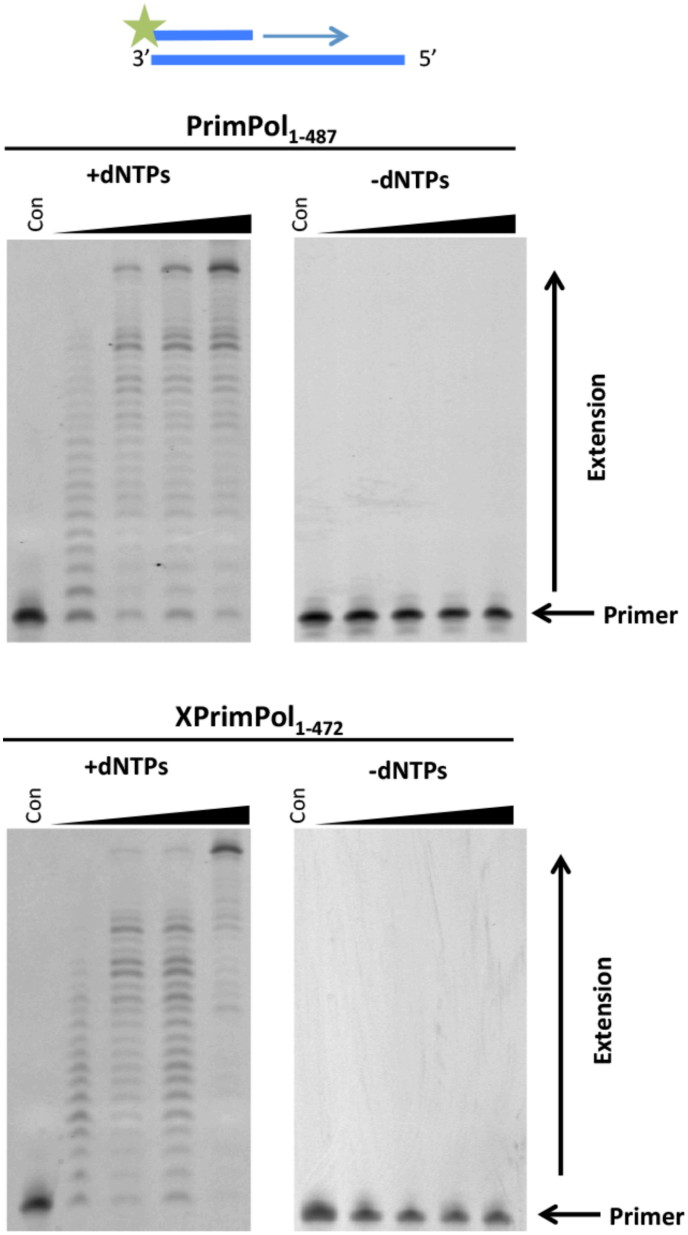


Figure 3.8. Polymerase activity of PrimPol$_{1-487}$ and XPrimPol$_{1-472}$. PrimPol$_{1-487}$ and XPrimPol$_{1-472}$ were incubated with dNTPs and substrate 1 from Table 2.6 at 1, 3, 5 and 30 minute time points alongside a control containing no protein. Both proteins are proficient at extending an undamaged oligonucleotide substrate using dNTPs. The proteins were also analysed in the absence of dNTPs and no activity was observed. The C-terminus of PrimPol, predicted to be unstructured, is not necessary for polymerase activity.

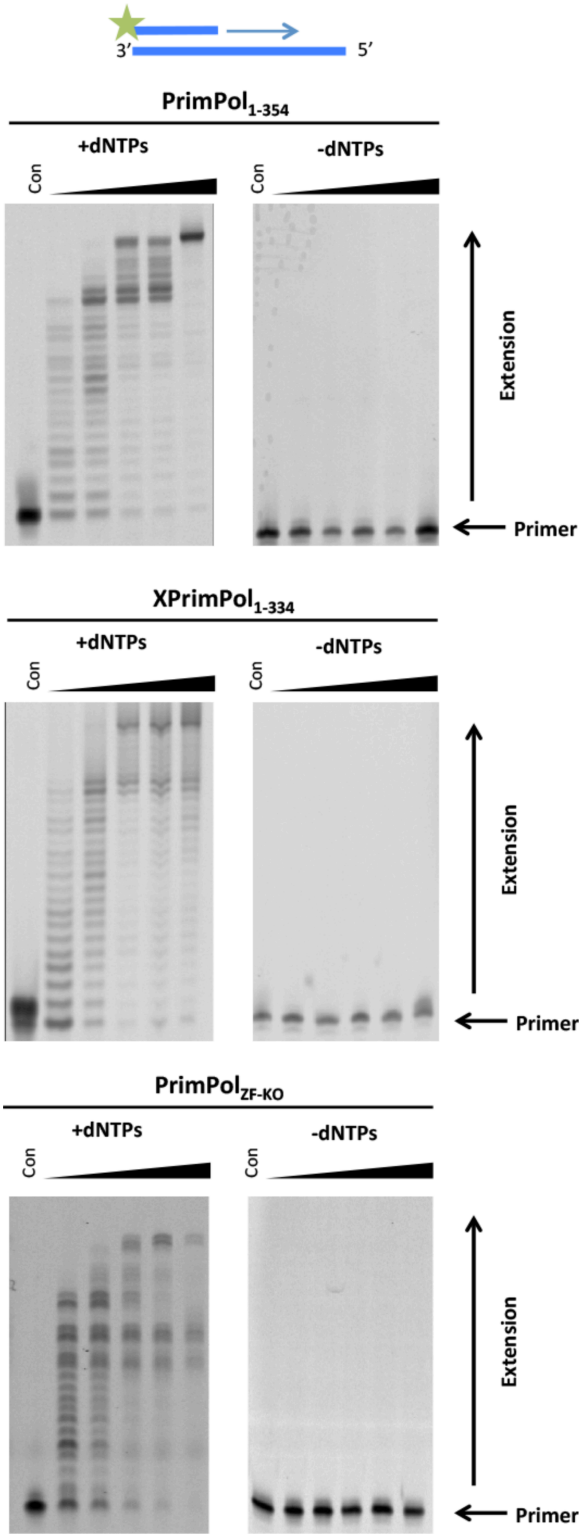


Figure 3.9. Polymerase activity of PrimPol$_{1-354}$ and XPrimPol$_{1-334}$, PrimPol$_{ZF-KO}$. PrimPol$_{1-487}$ and XPrimPol$_{1-472}$ were incubated with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. All three proteins are proficient at extending an undamaged oligonucleotide substrate using dNTPs. The proteins were also analysed in the absence of dNTPs and no activity was observed. The zinc finger of PrimPol is dispensable for PrimPol's polymerase activity.

3.6. PrimPol Cannot Utilise Ribonucleotides to Extend a Primer

As discussed in Section 1.7 many AEPs are able to produce and extend primers utilising both rNTPs and dNTPs. Surprisingly, PrimPol and its truncations $PrimPol_{1-487}$ and $PrimPol_{1-354}$ are unable to incorporate rNTPs when extending a DNA primer (Figure 3.10), despite being able to produce RNA primers (Bianchi et al., 2013). $XPrimPol_{1-334}$ was also assessed for its ribonucleotide incorporation activity to investigate whether this absent activity was limited to human PrimPol but it appears to be congruous across organisms (Figure 3.10). An example of an AEP with a preference for extending DNA with rNTPs is *M. paludicola* NHEJ pol (Bartlett et al., 2013), and this protein was used as a positive control to compare against the ribonucleotide incorporation activity of PrimPol; indeed, there is an extension of two rNTPs onto the DNA primer terminus as previously described (Bartlett et al., 2013) (Figure 3.10). It remains to be seen whether PrimPol can extend an RNA primer using rNTPs.

3.7. The Polymerase Domain of PrimPol can bind DNA Independently of the Zinc Finger

As PrimPol can extend primer DNA at a primer-template junction, we decided to test whether the AEP polymerase domain of PrimPol binds to single-stranded DNA, double-stranded DNA, or both. Electrophoretic mobility shift assays (EMSAs) were performed with increasing protein concentrations to analyse the DNA binding capabilities of the AEP polymerase domain of PrimPol. $PrimPol_{1-354}$ was found to bind to both ssDNA and dsDNA with approximately equal proficiency, binding was evident at 500 nM protein and a complete shift of substrate at 5 μM protein (Figure 3.11). The observed DNA binding specificity reflects the likely presence of two closely clustered binding sites allowing recognition of a primer-template junction, which is an essential requirement for all DNA polymerases. This relatively weak coordination of DNA substrate is also consistent with previously postulated biological functions of human PrimPol in TLS (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Rudd et al., 2013).

As a positive control for these EMSAs, a known DNA-binding protein *M. paludicola* Ku, which binds specifically to double-stranded DNA, was used to confirm that the DNA was shifted as a result of protein binding (Figure 3.11). To ensure that the DNA shift was not a result of molecular crowding caused by an increased concentration of protein, BSA was added to an overhanging DNA with both ssDNA and dsDNA; there is

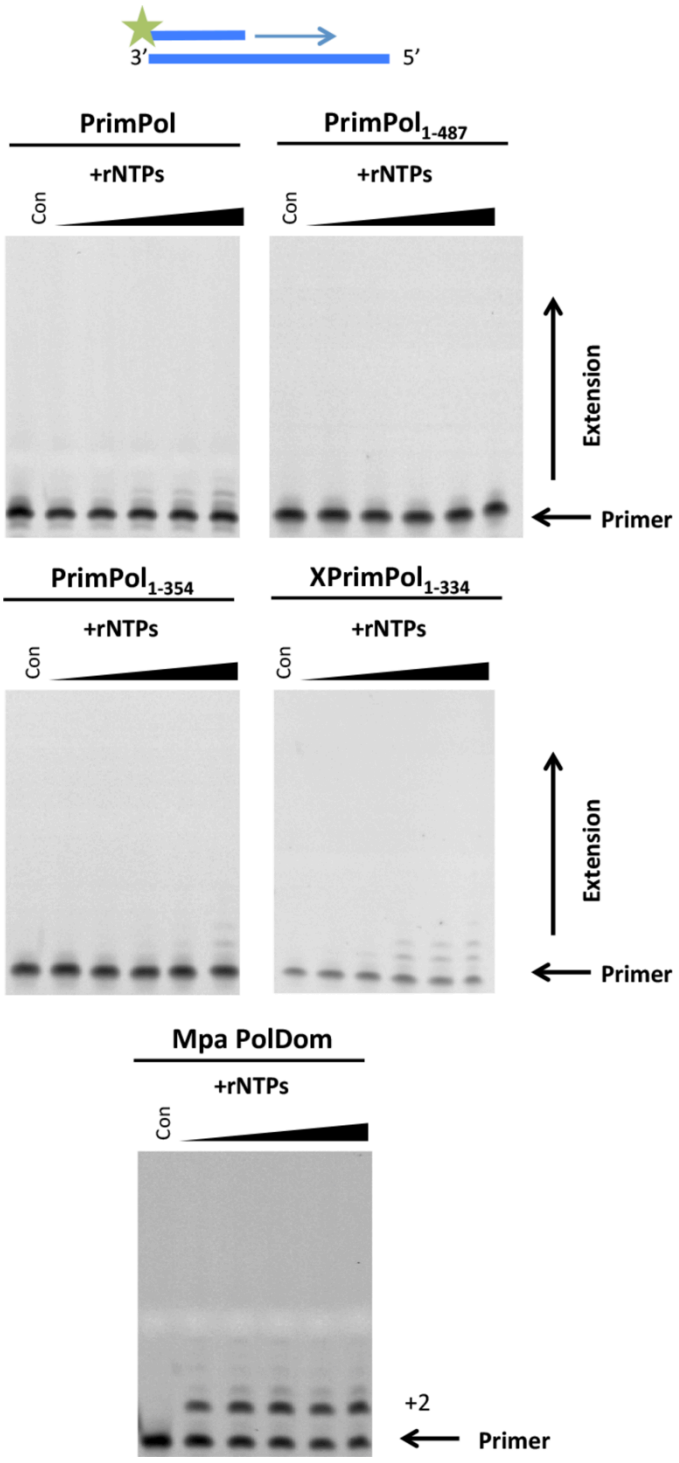


Figure 3.10. Assessment of the DNA-dependent RNA polymerase activity of PrimPol.
PrimPol, PrimPol$_{1-487}$, PrimPol$_{1-354}$ and XPrimPol$_{1-334}$ were incubated with rNTPs and an undamaged DNA substrate (substrate 1 from Table 2.6) at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. PrimPol is unable to extend a DNA primer using ribonucleotides opposite a DNA template. PolDom from M. paludicola (Mpa) was also analysed for its ribonucleotide incorporation activity as a positive control and incorporates two ribonucleotides.

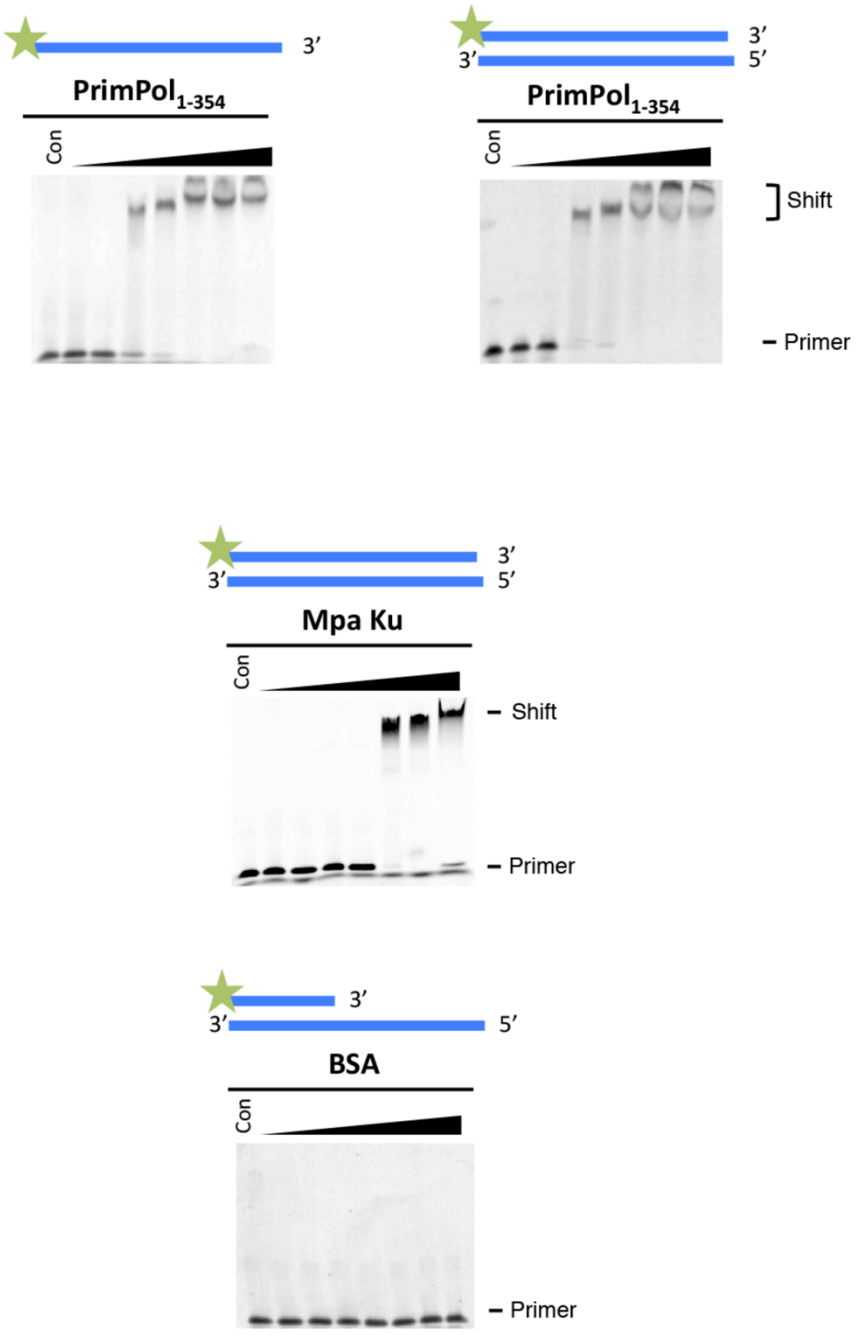


Figure 3.11. The polymerase domain of PrimPol binds both single stranded and double stranded DNA.
$PrimPol_{1-354}$ was analysed for DNA binding activity on 40 nM of both a single stranded and double stranded template (substrates 2 and 3 from Table 2.6, respectively) by EMSA at protein concentrations of 0, 0.05, 0.1, 0.5, 1.0, 5.0, 10.0 and 20.0 µM. PrimPol is able to bind both templates with comparable affinities as illustrated by the shift of the fluorescent oligonucleotide template on the polyacrylamide gel. Ku from M. paludicola (Mpa) was analysed as a positive control for DNA binding on the double stranded DNA template at concentrations of 0, 0.005, 0.01, 0.05, 0.1, 0.5, 1.0 and 5.0 µM. To ensure that the shift was not caused by molecular crowding by increasing protein concentration, protein shifts were analysed using an overhanging DNA substrate (substrate 1 from Table 2.6) and bovine serum albumin (BSA) at concentrations of 0, 0.05, 0.1, 0.5, 1.0, 5.0, 10.0 and 20.0 µM.

no shift when BSA was added at increasing concentrations to the fluorescent DNA (Figure 3.11). Upon testing full-length, wild-type PrimPol for its ability to bind single-stranded and double-stranded DNA, the DNA became shifted in a concentration-dependent manner but the shift caused a smearing on the gel with a significant portion of protein-DNA complexes remaining in the wells of the gel.

3.8. PrimPol's Zinc Finger Chelates a Zinc Ion

As described in Section 1.8, the C-terminal domain of human PrimPol contains a highly conserved UL52-like Cys-His-Cys-Cys (CHC2) motif, predicted to fold into a functional zinc finger motif (Figure 1.22). To establish if this is a *bona fide* zinc-binding motif, we produced C-terminal truncations in PrimPol containing either an intact ($PrimPol_{372-560}$) or mutated ($PrimPol_{372-560/ZF-KO}$) zinc finger motif. $PrimPol_{372-560/ZF-KO}$ has the first two invariable residues of the zinc finger motif mutated to alanine. The proteins were gel filtered in the absence of zinc and concentrations of zinc were then measured using inductively coupled plasma mass spectrometry (ICP-MS) (Table 3.1). ICP-MS is used for elemental analysis, typically that of metal ions. Samples are dried, vapourised, atomised and ionised in plasma composed of argon ions at a temperature of over 6000 K. Ions are filtered by their mass/charge ratio using a quadrupole mass filter and, finally, analysed by mass spectrometry. The ICP-MS analysis revealed that the concentration of the zinc in the UL52-like domain represented an occupancy of 80.38%, suggesting that a small fraction of the protein is either unfolded or has another divalent metal ion chelated. This is a much higher zinc concentration than evident for the AEP BcMCM (11%) (Samuels et al., 2009); however, zinc concentration for BcMCM was determined following purification in the absence of zinc and subsequent dialysis into another buffer that lacked zinc, allowing for the titration out of zinc ions at a number of steps.

In contrast, the zinc finger knockout ($PrimPol_{ZF-KO}$) had an occupancy of ~3.90%, suggesting that the first two cysteine and histidine residues of this motif are required for the coordination of zinc; the presence of a minute level of zinc suggests that either the protein folds into a structure that has some weak zinc-binding activity or that some of the zinc was transferred into the gel filtration buffer from earlier in the purification. We also measured the zinc occupancy in the polymerase domain ($PrimPol_{1-354}$) after purification in the absence of zinc and the occupancy in this case was ~1.57%. The detection of low levels of zinc in the polymerase domain is to be expected as it contains

Protein Construct	Protein Concentration	Zinc Concentration	Occupancy
PrimPol$_{372-560}$	80.0	64.34 ± 1.29	80.38%
PrimPol$_{372-560/ZF-KO}$	30.0	1.17 ± 0.05	3.90%
PrimPol$_{1-354}$	32.5	0.51 ± 0.023	1.57%

Table 3.1. Zinc occupancy of PrimPol domains.
The PrimPol zinc finger domain (PrimPol$_{372-560}$) and zinc finger domain knockout (PrimPol$_{372-560/ZF-KO}$) were gel filtered in the absence of zinc to remove any zinc that was not chelated by protein. They were subsequently tested for zinc concentration by inductively coupled plasma mass spectrometry (ICP-MS). The zinc finger domain exhibits zinc occupancy of 80.43%, whereas the zinc finger knockout only has occupancy of 3.9%, suggesting that the zinc chelation requires the intact amino acids that form the zinc finger motif (Cys-His-Cys-Cys, not Ala-Ala-Cys-Cys). The 1-354 polymerase domain was purified in the absence of zinc and was also subsequently tested for zinc occupancy, but only had 1.57% occupancy.

divalent cation-binding motifs. These motifs generally bind magnesium or manganese, which have respective ionic radii of 86 and 81 pm, but could conceivably bind zinc, with a comparable ionic radius of 88 pm (Shannon, 1976).

3.9. The Zinc Finger Domain Binds Single-Stranded DNA

Having established that the putative zinc finger does indeed chelate a zinc ion, we aimed to determine whether it was capable of binding DNA. The zinc finger domain is presumed to be a DNA-recognition domain, in addition to contributing to its primase activity, as previous studies demonstrated altered DNA binding activity associated with mutations in the related zinc finger domain of UL52 (Biswas and S. K. Weller, 1999; Y. Chen et al., 2007). EMSAs were performed with increasing protein concentrations to analyse the DNA binding capabilities of the zinc finger of PrimPol, controls for this experiment are described in Section 3.7.

In contrast with the AEP polymerase domain, PrimPol's zinc finger domain was unable to bind dsDNA but did show specific binding to ssDNA (Figure 3.12), albeit with a low apparent dissociation constant. This intriguing result provides experimental evidence supporting the potential importance of the zinc finger module in recognition/coordination of the single-stranded template DNA upstream of the primer-template junction. To determine if it was indeed the zinc finger that performed DNA binding, we assayed a zinc finger knockout mutant domain ($PrimPol_{372\text{-}560/ZF\text{-}KO}$) for DNA binding activity (Figure 3.12) and observed an abolishment of binding with this mutant, confirming the importance of the zinc finger motif for nucleic acid recognition.

3.10. An Intact Zinc Finger is Required for Primase Activity

It was previously shown that PrimPol has DNA primase activity *in vitro* (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Wan et al., 2013). Although the zinc finger of PrimPol is not required for its polymerase activity, it may have been a requirement for primase activity. Primase assays were carried out in collaboration with my colleague Dr. Stanislaw Jozwiakowski. Primers were synthesised *de novo* opposite homopolymeric templates that were 5'-biotinylated using either ribonucleotides or deoxyribonucleotides. Primers were extended using 2 U Klenow Taq supplemented with a fluorescence nucleotide analogue to allow for the fluorescent labelling of primers. Reactions were quenched and DNA bound to streptavidin-coated beads and excess

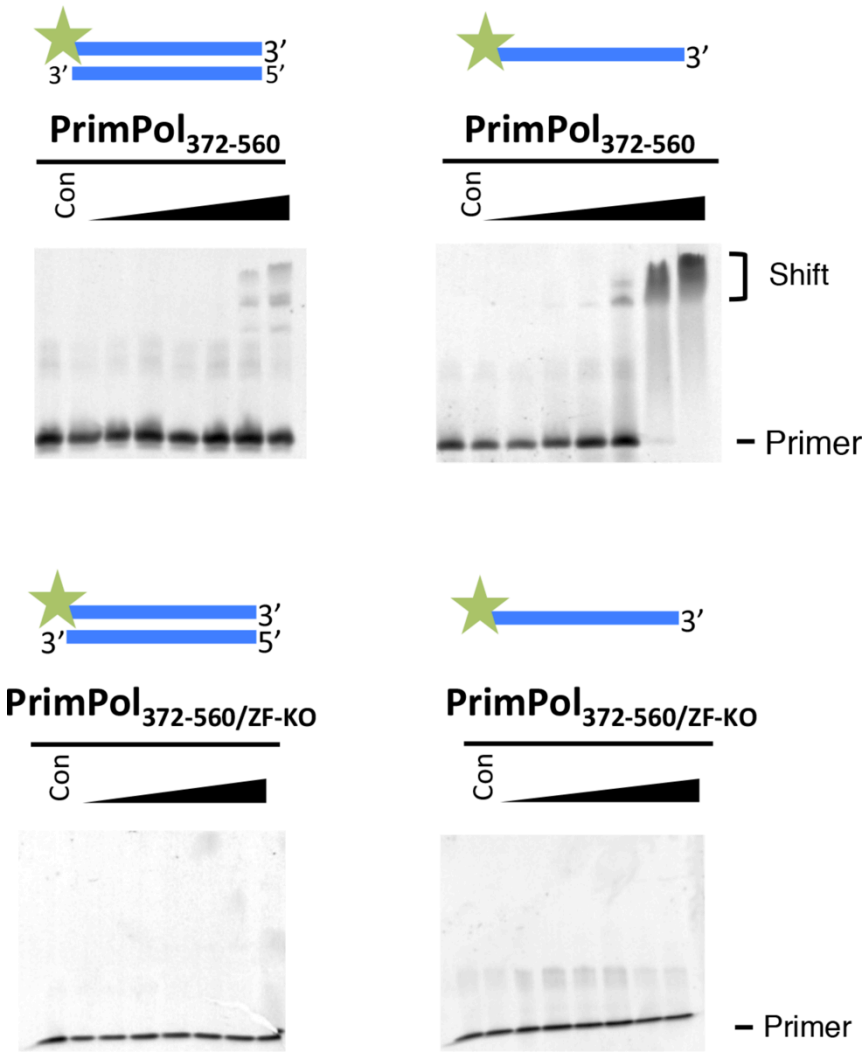


Figure 3.12. The zinc finger of PrimPol binds single stranded DNA.
$PrimPol_{372-560}$ and $PrimPol_{372-560/ZF-KO}$ were analysed for DNA binding activity on 40 nM of both single stranded and double stranded templates (substrates 2 and 3 from Table 2.6, respectively) by EMSA at protein concentrations of 0, 0.05, 0.1, 0.5, 1.0, 5.0, 10.0 and 20.0 µM. $PrimPol_{372-560}$ is only capable of binding ssDNA, with the shift on dsDNA much less pronounced. $PrimPol_{372-560/ZF-KO}$ is unable to bind either single-stranded or double-stranded DNA. An intact zinc finger is required for PrimPol C-terminal binding of ssDNA.

nucleotides were washed prior to resolution on a polyacrylamide gel and detection by fluorescence scanner (Keen et al., 2014).

The zinc finger motif of the related AEP UL52 is an absolute requirement for primase activity (Biswas and S. K. Weller, 1999; Y. Chen et al., 2005). As previously observed (Bianchi et al., 2013; Rudd et al., 2013), wild-type PrimPol exhibited primase activity on a 60-mer poly-dT ssDNA template, however $PrimPol_{ZF\text{-}KO}$ showed no detectable primase activity (Figure 3.13). Furthermore, we analysed the primase activity of two deletion mutants, $PrimPol_{1\text{-}487}$ and $PrimPol_{1\text{-}354}$, which were constructed to determine whether elimination of primase activity resulted from a structural change brought about by the point mutations or whether the zinc-finger domain is necessary for primase activity. As expected, the $PrimPol_{1\text{-}354}$ deletion mutant lacking the entire zinc finger showed no detectable primase activity (Figure 3.13). However, $PrimPol_{1\text{-}487}$, containing the entire zinc-finger domain but lacking the smaller C-terminal portion was an active DNA primase (Figure 3.13). Notably, the observed primase activity of $PrimPol_{1\text{-}487}$ was lower than the activity observed for the parental enzyme, as this truncation possibly results in partial loss of the zinc finger domain structure and functionality.

Together, our data provide experimental evidence that the zinc-finger domain, but not the disordered C-terminal region proximal to this domain, is essential for the primase activity of human PrimPol. PrimPol without an intact zinc finger can, therefore, be used as a “separation of function” mutant to help determine which catalytic functions, either the primase activity or polymerase activity of PrimPol, are required *in vivo*.

3.11. PrimPol’s Polymerase Domain is Essential for the Maintenance of Normal Replication Fork Rate in Unperturbed Cells

It was previously shown by our laboratory that avian DT40 cells that had their PrimPol knocked out ($PrimPol^{-/-}$) showed a significant reduction in replication fork progression rate (Bianchi et al., 2013). Given the biochemical data established in this chapter, we aimed to evaluate the requirement for the zinc finger for replication fork rate progression. It was previously shown that stably expressing wild-type PrimPol in these $PrimPol^{-/-}$ DT40 cells would successfully restore near wild-type replication fork levels rates (Bianchi et al., 2013). This *in vivo* analysis was carried out in collaboration with

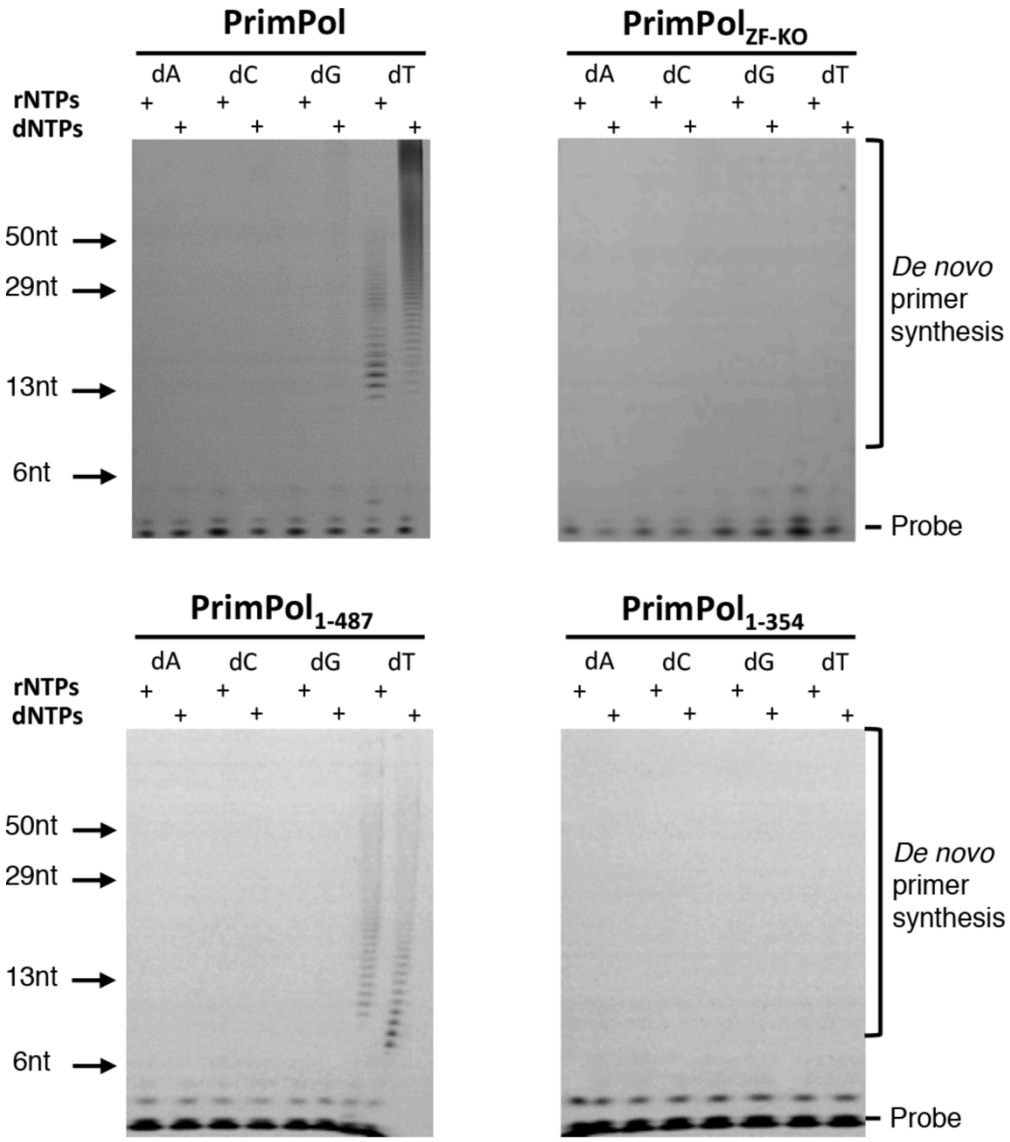


Figure 3.13. DNA primase activity of PrimPol.
PrimPol, PrimPol$_{ZF-KO}$, PrimPol$_{1-487}$ and PrimPol$_{1-354}$ were all analysed for DNA-dependent primase activity on homopolymer substrates (substrates 4-7 in Table 2.6) using either rNTPs or dNTPs. PrimPol and PrimPol$_{1-487}$ both have de novo primer production activity using both rNTPs and dNTPs on poly(dT) substrates, with a preference for utilisation dNTPs. Neither PrimPol$_{ZF-KO}$ nor PrimPol$_{1-354}$ exhibit *de novo* primase activity. The experiments shown here were carried out by Dr. Stanislaw Jozwiakowski.

my colleague Dr. Laura Bailey. Replication fork rates were measured by DNA fibre analysis, which involves the pulse labelling of exponentially growing DT40 cells using CldU for 20 minutes, followed by IdU for 20 minutes. Labelled DNA was then detected using anti-CldU and anti-IdU antibodies (Edmunds et al., 2008). Replication rates were then calculated as described in Section 2.5.5.

We confirmed that normal replication fork rates are restored by adding back human PrimPol (Figure 3.14); however, this replication fork defect is not corrected by adding back the catalytically dead $PrimPol_{AXA}$ mutant (Motif I DxE→AxA; D114A, E116A mutant) (Figure 3.14). Next, we measured the replication fork speed in $PrimPol_{1-354}$ and $PrimPol_{ZF-KO}$ complemented clones. We observe that both proteins were able to restore replication fork rates to near wild-type levels (Figure 3.14). This suggests that only the polymerase activity of PrimPol, not the primase activity, is required for the maintenance of fork progression in unperturbed replication. $PrimPol_{1-354}$ complemented cells show even greater average replication fork rates compared to wild type cells (Figure 3.14), in accordance with its moderately superior polymerase activity *in vitro* (Section 3.5). These *in vivo* findings are in accordance with other studies (Mourón et al., 2013).

3.12. Towards the Structural Characterisation of Human PrimPol

To gain a better understanding of the amino acids required for catalysis and how PrimPol carries out its catalytic activities, we aimed to crystallise the AEP polymerase domain of PrimPol. The catalytic domain alone was selected for attempting to resolve the structure of PrimPol as previous investigations, including those in our own laboratory, had often been successful in the first instance in crystallising the catalytic subunit before attempting to crystallise with accessory peptides (Augustin et al., 2001; Brissett et al., 2007; N. Ito et al., 2003). The polymerase domain $PrimPol_{1-354}$ was chosen to attempt crystallisation but following extensive screening I was unable to produce any crystals of any suitable size for X-ray diffraction and those small crystals that did develop were unfortunately not reproducible.

As the first predicted α helix of human PrimPol is not conserved between human and *X. tropicalis* (Figure 3.1, Appendix I), we decided to remove this first helix and attempt crystallisation again. The first 24 amino acids of human PrimPol were removed from the polymerase domain and the resulting protein was purified (Figure 3.15A). Similar to

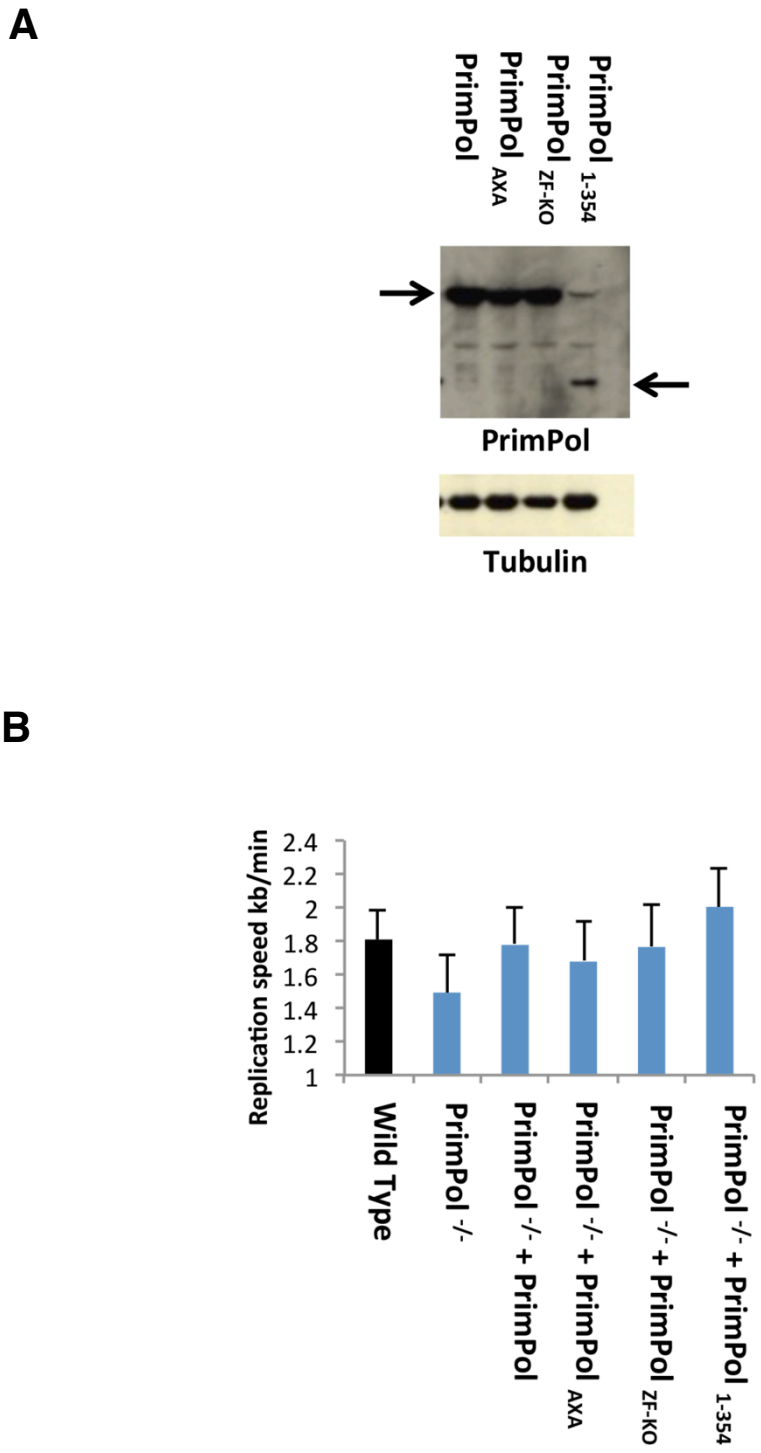


Figure 3.14. PrimPol polymerase activity is required for maintaining replication fork rate *in vivo.*

(A) *PrimPol*$^{-/-}$ DT40 cells were complemented with wild-type, AXA (D114A/E116A), 1-354, and ZF-KO variants of PrimPol and their *in vivo* expression confirmed by western blot. **(B)** Unperturbed replication fork rates were analysed by DNA fibre analysis using CldU and IdU labels for all the complemented cell types. The data represent the mean of three or more experiments with error bars showing the mean standard deviation. Experiments shown here carried out by Dr. Laura Bailey.

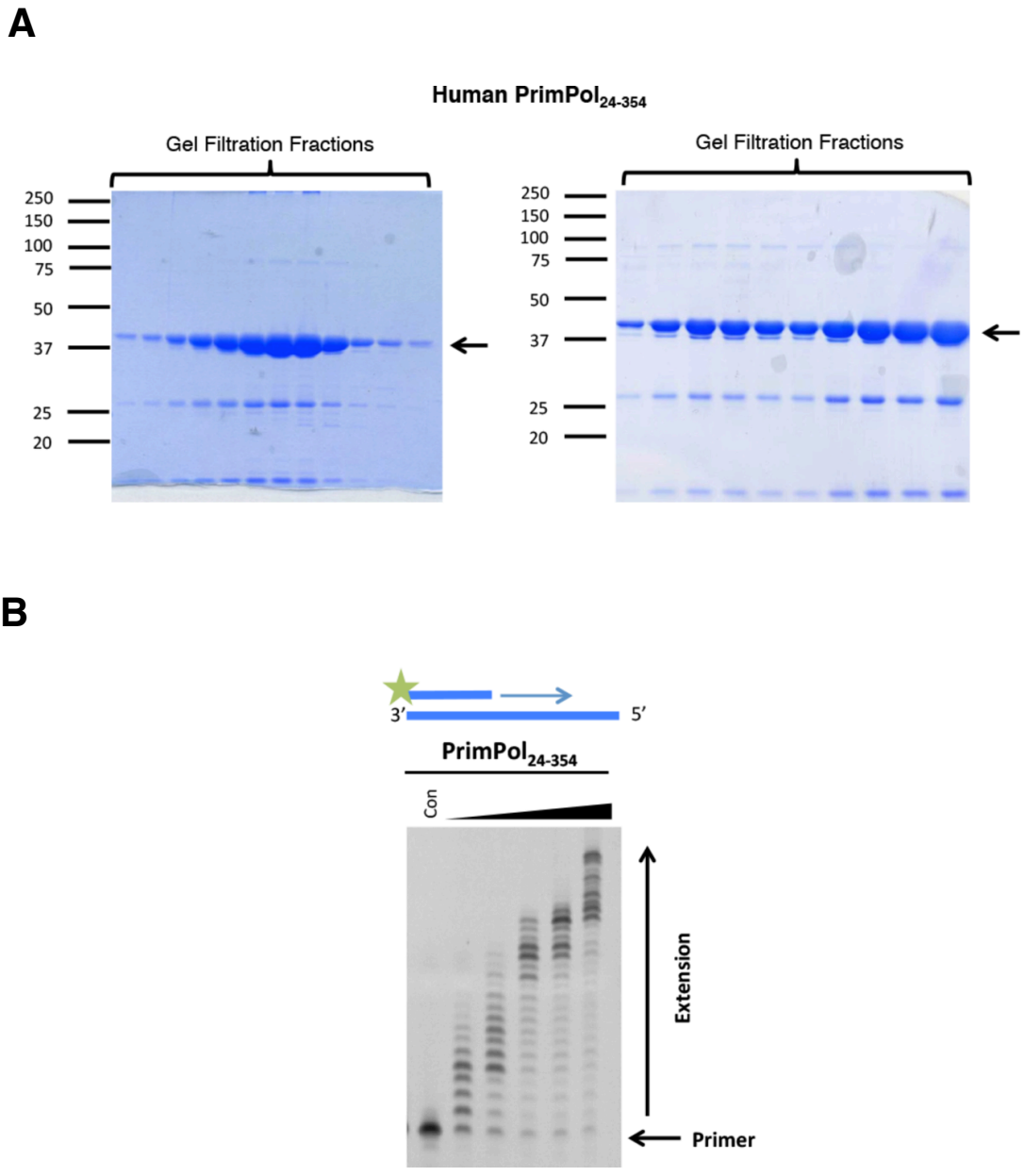


Figure 3.15. Purification and polymerase activity of PrimPol$_{24-354}$.
(A) SDS-PAGE analysis of PrimPol$_{24-354}$ showing samples from the protein-containing fractions from the size exclusion chromatography of PrimPol$_{24-354}$ (40.3 kDa). Breakdown products are evident and these products are analysed by mass spectrometry in Section 3.12.1. The second gel (top right) represents SDS-PAGE analysis of the size exclusion chromatography fraction samples upon expression and purification of PrimPol$_{24-354}$ in the presence of increased benzamidine and PSMF protease inhibitors showing the persistence of PrimPol breakdown products. **(B)** PrimPol$_{24-354}$ was incubated with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points, along with a control containing no protein. PrimPol$_{24-354}$ is proficient at extending an undamaged oligonucleotide substrate using dNTPs.

$PrimPol_{1-354}$ (Figure 3.4) and $XPrimPol_{1-334}$ (Figure 3.6) two smaller protein bands are evident on the $PrimPol_{24-354}$ purification SDS-PAGE gel. As it was initially suspected that these breakdown bands were caused by proteases from purification, I increased the protease inhibitor concentration two-fold and attempted to re-purify the protein (Figure 3.15A). However, these protein degradation bands were still present, suggesting either that these bands represented peptides that were not $PrimPol_{24-354}$ but interacted with PrimPol during purification, peptides that have the same chromatography column-binding profiles as $PrimPol_{24-354}$, or $PrimPol_{24-354}$ having an exposed site that is susceptible to nicks in the protein through a protease inhibitor-resistant means. These degradation bands are analysed in section 3.12.1. $PrimPol_{24-354}$ was confirmed to retain its polymerase activity upon removal of this putative first α helix (Figure 3.15B). Unfortunately, following more extensive screening with $PrimPol_{24-354}$, the same issues were experienced as with $PrimPol_{1-354}$ in attempted generation of protein crystals.

3.12.1. Identification of PrimPol Degradation Products

As described in the section above, when purifying the human and *X. tropicalis* polymerase domains, two smaller peptides are evident on the SDS-PAGE analysis gels. To identify whether these proteins are PrimPol and, if so, where the protein is being nicked, these protein bands were analysed by in-gel protease digestion and mass spectrometry (see Section 2.4.2). The resulting short peptides identified the smaller proteins observed on the gels to indeed be truncated forms of PrimPol's polymerase domain.

The largest protein band observed on the gel was confirmed to be PrimPol and peptides covered the length of the protein (Figure 3.16A). The middle band at around 30 kDa on the protein did not produce any peptides beyond amino acid 271, suggesting that the protein degrades at residue 271 or somewhere slightly further towards the C-terminus (Figure 3.16B). The lower band's digestion products ranged from amino acids 170-344 (result not shown); as a single protein product this size represents a protein around 20 kDa and the band in the gel suggests this protein is much smaller than that, it is likely that this band actually represents the amino acids C-terminal of this truncation at around residue 271 and other peptides present are contaminating small peptides one might expect to be present in small quantities towards the bottom of the electrophoresed polyacrylamide gel.

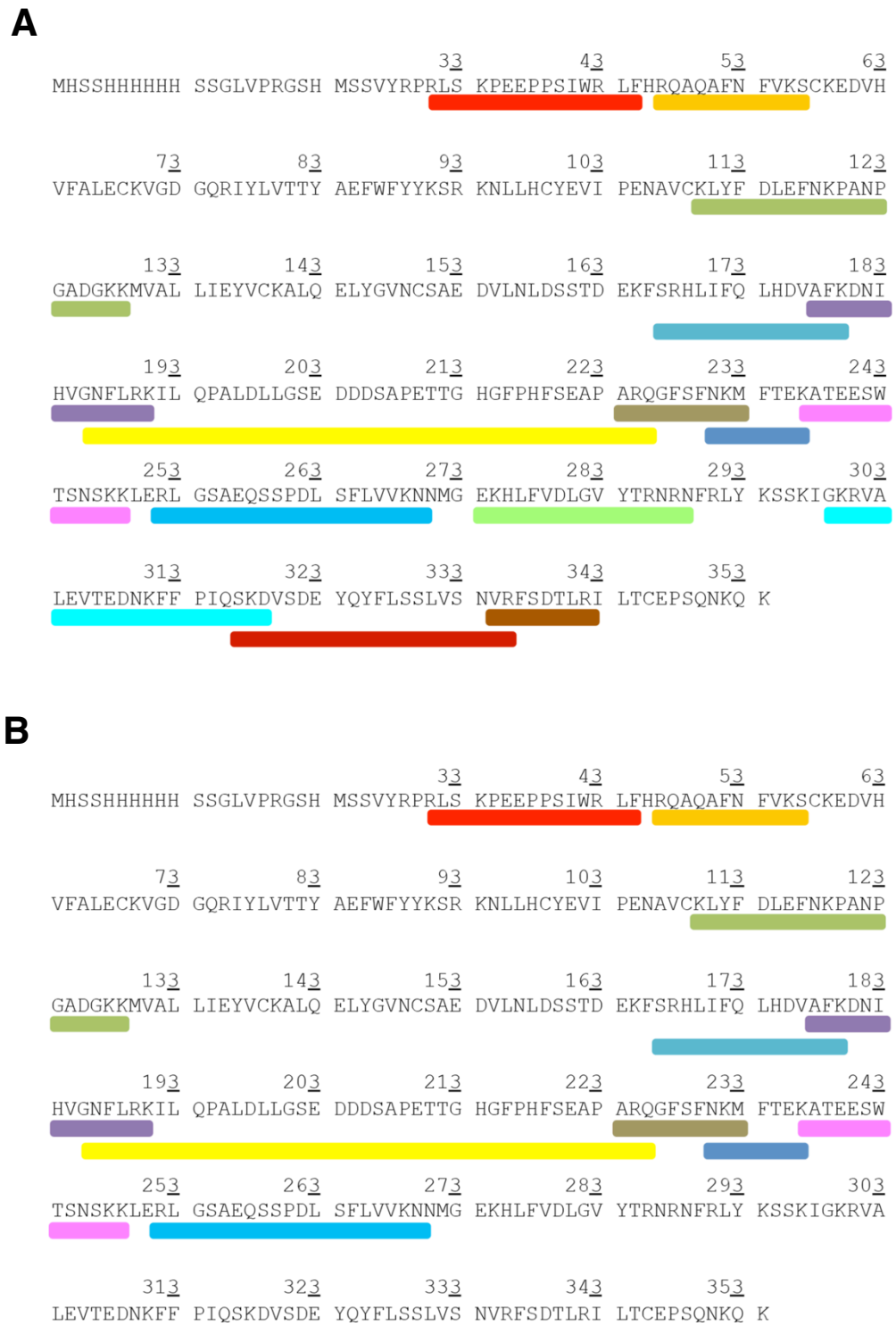


Figure 3.16. Identification of PrimPol$_{24\text{-}354}$ breakdown products.
In-gel digestion and mass spectrometry of breakdown products of PrimPol$_{24\text{-}354}$ compared with the major, non-truncated protein band. Protein fragments are depicted as block colours beneath corresponding amino acids. Numbers correspond to the amino acid numbers in human PrimPol. **(A)** The protein fragments from the major band of purified PrimPol$_{24\text{-}354}$ range the length of the protein as expected. **(B)** The top breakdown band protein fragments from in-gel digestion range up to amino acid 271 suggesting that the protein is truncated at some point after residue 271.

3.12.2. Hydrogen-Deuterium Exchange Reveals the Unstructured Regions of Human PrimPol

We aimed to identify more of the unstructured regions that were hindering crystallisation by probing human PrimPol through hydrogen-deuterium exchange. This work was carried out in collaboration with the Skehel laboratory at the MRC Laboratory of Molecular Biology in Cambridge. The principle of this approach involves the analysis of peptides following a period of exchange between hydrogen and its heavier isotope deuterium ($^{2}_{1}\mathrm{H}$) in a solution of D_2O. Following this exchange, peptides are analysed for their level of exchange by mass spectrometry; high exchange suggests a more exposed, external region of the protein, whilst less exchange suggests more of a central, internal region of the protein.

The hydrogen-deuterium exchange profile of PrimPol is shown in Figure 3.17. As expected, the core, conserved amino acids show less exchange with D114 and E116 of motif I, S167 and H169 of motif II, and D280 of motif III all found in low-exchange regions of the protein. There is a high-exchange region in the catalytic domain from around amino acid 220. Intriguingly, part of this high-exchange region is absent from both *X. tropicalis* and *Mus musculus* (Mouse) PrimPol (Figure 3.17), suggesting perhaps this unstructured region is unnecessary for catalysis and may be mobile, thus impeding crystallisation. The hydrogen-deuterium exchange data indicates that, at the beginning of the protein, the amino acids on the N-terminus of residue 41 are exposed and flexible as demonstrated by their propensity for hydrogen-deuterium exchange. At the C-terminus of the polymerase domain, there appears to be exchange as early as amino acid 291. This suggests the polymerase domain could be further truncated to construct a more minimal and, consequently, stable polymerase domain for structural studies.

3.12.3. A Minimal PrimPol AEP Polymerase Domain

As our mass spectrometry data suggested that the protein had a fragile site at around amino acid 271 or slightly higher (Section 3.12.1), the protein was truncated on the N-terminal side of a predicted beta sheet at amino acid 276 (Appendix I). This resultant recombinant protein was, as expected, inactive (Figure 3.18) as it precludes the third essential motif (D at position 280) for catalysis. Concurrently, we also produced a construct of amino acids 39-276. Amino acid 39 was selected as it offers a 2 amino

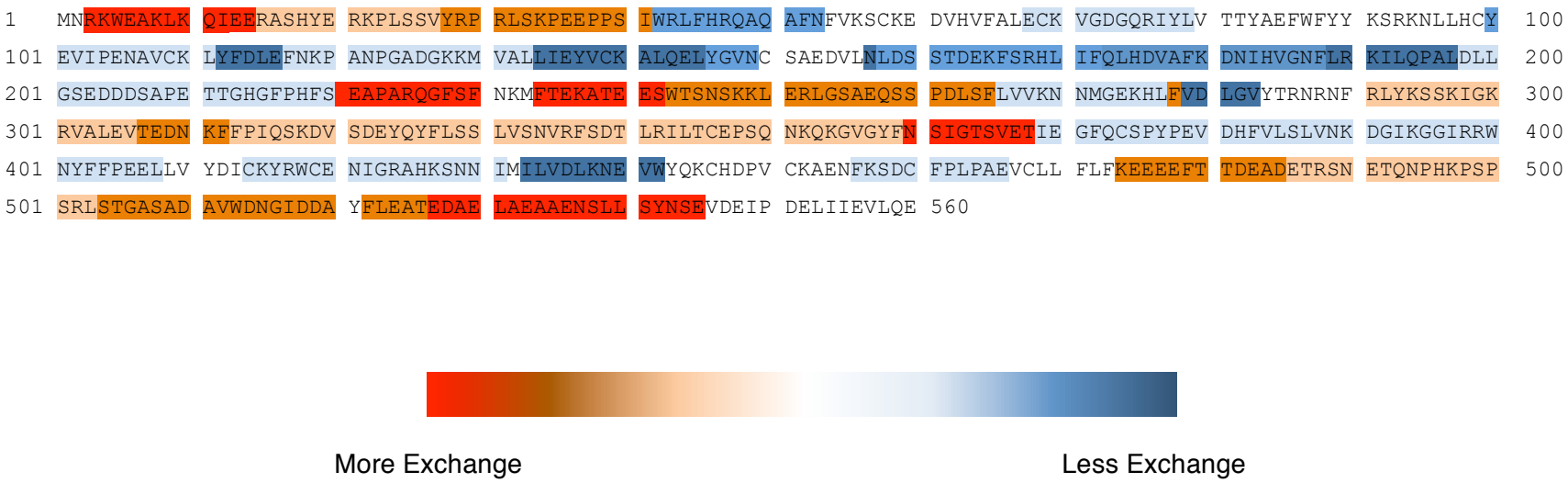


Figure 3.17. Hydrogen-Deuterium Exchange of PrimPol.
Protein areas of high isotopic exchange are indicated in red and those areas of lower isotopic exchange are indicated in blue. Low exchange is indicative of the more protected, central regions of protein, whilst high exchange is indicative of more exposed, external regions of protein. These experiments were carried out in the laboratory of J. Mark Skehel.

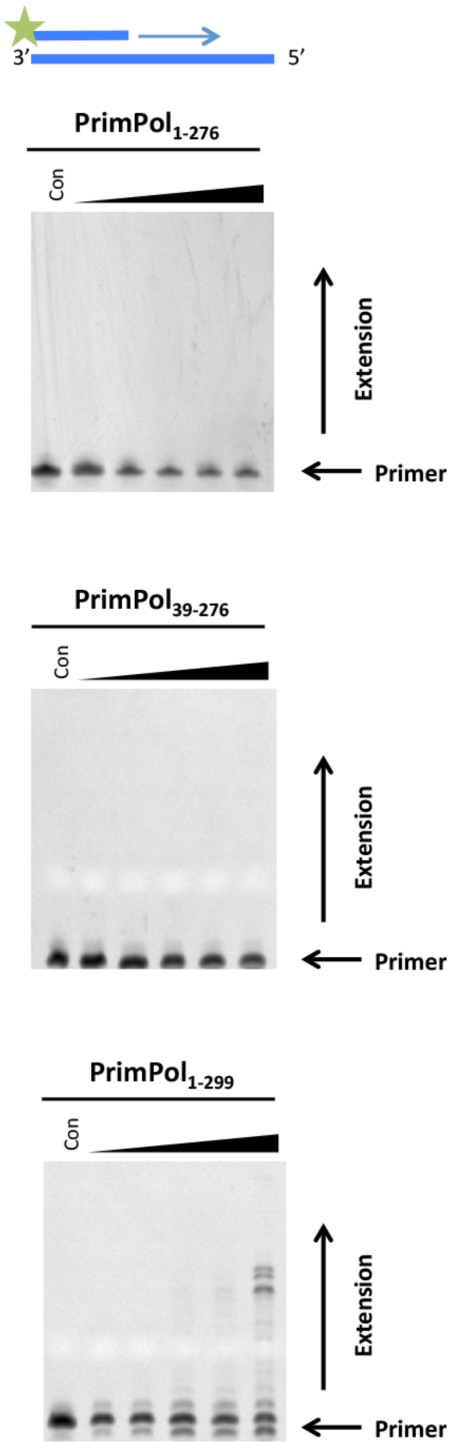


Figure 3.18. Polymerase activity of $PrimPol_{1-276}$, $PrimPol_{39-276}$ and $PrimPol_{1-299}$. $PrimPol_{1-276}$,
$PrimPol_{39-276}$ and $PrimPol_{1-299}$ were constructed and purified as indicated in Tables 2.3-2.6. Each variant was incubated with dNTPs and DNA substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points, along with a control containing no protein. Neither $PrimPol_{1-276}$ nor $PrimPol_{39-276}$ had any polymerase activity. $PrimPol_{1-299}$ had some residual polymerase activity but also showed some exonuclease activity, suggesting the presence of contaminating proteins.

acid buffer to the putative highly structured amino acids that begin at position 41, as predicted from both our secondary structure prediction and our hydrogen-deuterium exchange data; the protein would also then begin with a proline to aid in demarcating the desired the protein and the N-terminal tag. $PrimPol_{39-276}$, too, was inactive as expected (Figure 3.18). The third construct we tested was $PrimPol_{1-299}$. 1-299 was chosen as the hydrogen-deuterium data determined that the region C-terminal of residue 291 was more exposed; additionally, the amino acids C-terminal of 299 are less well conserved between human and *X. tropicalis* than the amino acids to the N-terminal side of 299. $PrimPol_{1-299}$, however, showed a significant reduction in activity (Figure 3.18). There is only evidence of extension at 60 minutes and it is limited. As $PrimPol_{1-299}$ was only purified by affinity chromatography it cannot also be ruled out that the meagre activity that is present may be the result of contaminating proteins, which could also account for the small amount of exonuclease activity also present at increasing time points. In either case, this protein exhibits very little activity in comparison to $PrimPol_{1-354}$ or $PrimPol_{24-354}$.

In addition to these three constructs, $PrimPol_{39-354}$ was also cloned. $PrimPol_{39-354}$ purified and tested for activity (Figure 3.19). This $PrimPol_{39-354}$ construct had an unusual activity, whilst it was reduced relative to $PrimPol_{1-354}$ or $PrimPol_{24-354}$ it also had peculiar signature of primer extension products, only extending primers around halfway up the length of the template. Despite this altered activity, we went ahead with screening for crystallography products; the first human Y-family polymerase structure was resolved on a truncated form of pol κ that had reduced polymerase activity but whose polymerase domain remained intact (Uljon et al., 2004). Again though, following extensive crystal screening, this protein remained recalcitrant to producing crystals suitable for X-ray diffraction.

3.13. Towards the Structural Characterisation of Mouse PrimPol

To overcome the intractability of the human protein, we chose to study the *M. musculus* (Mouse) PrimPol, which has a highly similar sequence to human PrimPol. However, absent from mouse PrimPol's AEP polymerase domain are a number of amino acids (Figure 3.1) that are present in the human PrimPol but are predicted to be unstructured by hydrogen-deuterium exchange and could represent an unnecessary mobile loop element that posed a hindrance to crystallography trials. Given this, structural studies were continued using the mouse PrimPol as a potential candidate for crystallography

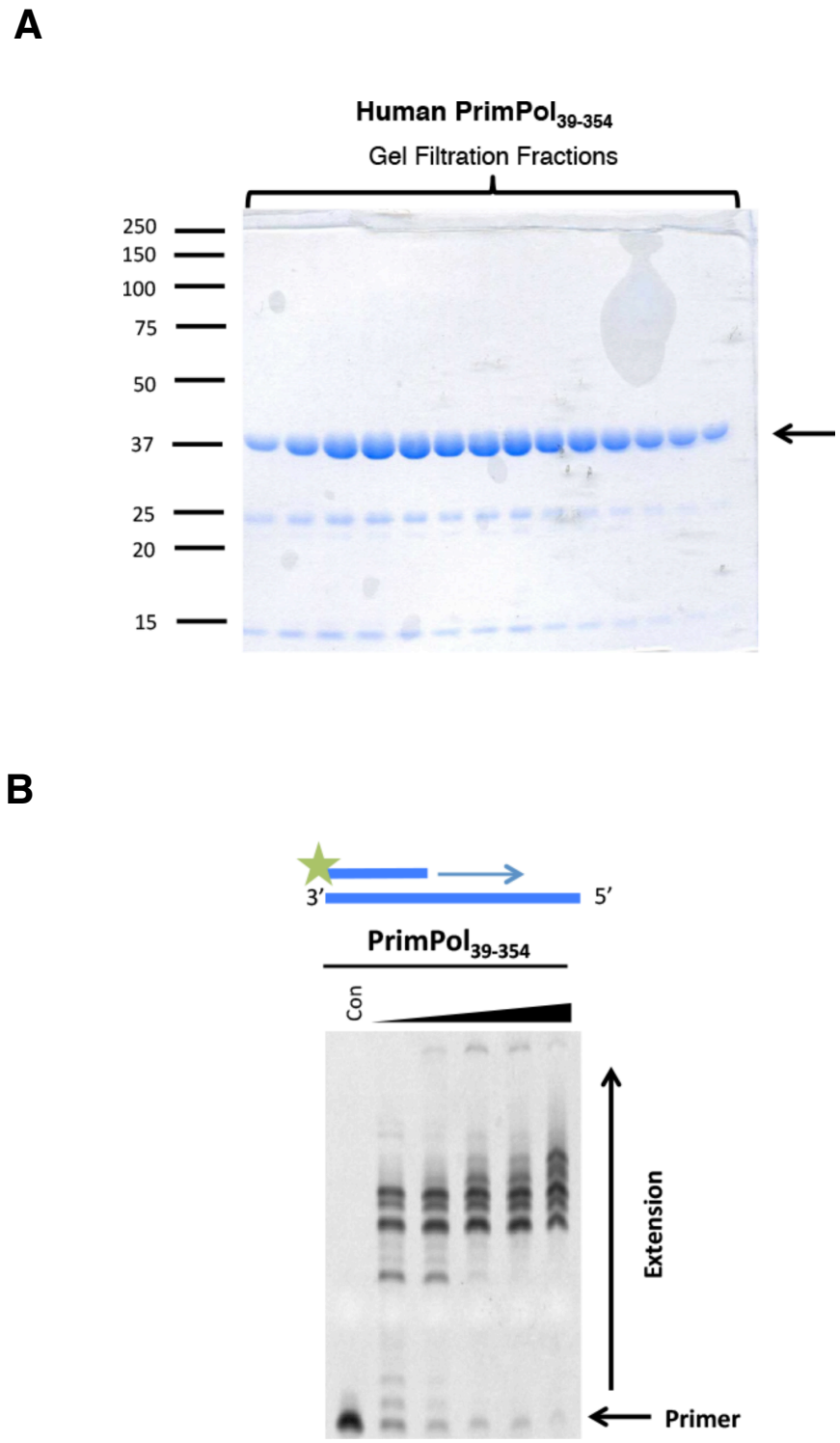


Figure 3.19. Purification and polymerase activity of PrimPol$_{39-354}$.
(A) SDS-PAGE analysis of PrimPol$_{39-354}$ showing samples from the protein-containing fractions from the size exclusion chromatography of PrimPol$_{39-354}$ (38.6 kDa). PrimPol breakdown products of remain evident. **(B)** PrimPol$_{39-354}$ was incubated with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points, along with a control containing no protein. PrimPol$_{39-354}$ has a strange signature of activity, appearing to stall at various points when extending the substrate primer; there is little evidence of full extension, even after 60 minutes.

trials. The equivalent construct of the human PrimPol AEP polymerase domain was initially constructed; in mouse PrimPol this was amino acids 1-338 ($MPrimPol_{1-338}$). This protein was purified and, unlike the human or *X. tropicalis* polymerase domains, there is no evidence of any degradation products on the gel (Figure 3.20A). The reason for this lack of protein breakdown, or rather the presence of this breakdown in human and *X. tropicalis* purifications, remains elusive.

$MPrimPol_{1-338}$ was shown by primer extension assays to be an active polymerase (Figure 3.20B). The protein was then screened for protein crystal production; however, in keeping with the human PrimPol polymerase domain, this domain was recalcitrant to crystallisation. The protein was then screened for crystals in the presence of a DNA template at a protein-DNA molar ratio of 1:1.5, and 4mM $MgCl_2$. The DNA template (substrate C1, Table 2.6) was chosen based on the DNA template for the crystallisation of pol η (Biertümpfel et al., 2010), with the length altered to a minimal length that the polymerase domain of PrimPol would still extend from (results not shown). This screening garnered reproducible crystals, in a diammonium phosphate condition, that were large enough in size to potentially undergo X-ray diffraction (Figure 3.21A). However, a simple crush test confirmed that these crystals were extremely unlikely to contain protein and were more likely to be salt crystals.

Despite a pre-crystallisation test (PCT) suggesting that the optimal concentration for screening $MPrimPol_{1-338}$ was 8.5 mg mL^{-1}, this concentration of protein gave a high proportion of sitting drops that exhibited large amounts of precipitation. As such, the protein concentration was lowered to 7 mg mL^{-1} and the DNA concentration adjusted to maintain a 1:1.5 molar ratio and the protein was re-screened for crystals. Crystals developed in two similar conditions containing similar reservoir buffers. These buffers were supplemented with 20% ethylene glycol as a cryoprotectant and X-ray diffraction data was collected and a diffuse pattern suggests that there is no ice crystal formation (representative example in Figure 3.21B). The crystals from the two conditions were then both analysed by X-ray diffraction (Figure 3.21C-D). Both diffraction patterns are indicative of salt crystals, high intensity spots at mid-high resolution; the increased number of spots in Figure 3.21D suggest that the salt crystals in this condition may contain some low level of protein.

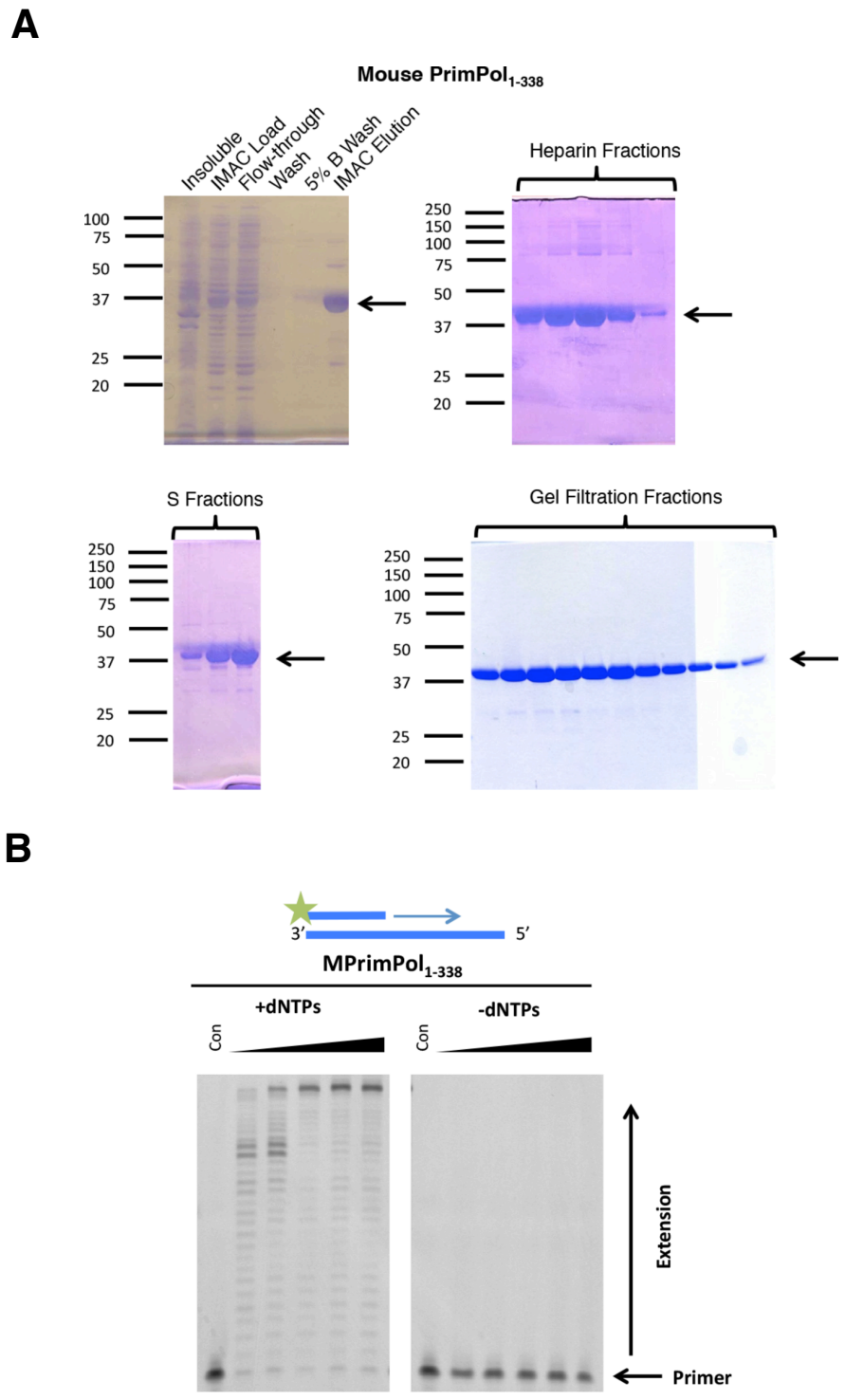


Figure 3.20. Purification and polymerase activity of MPrimPol$_{1\text{-}338}$.
(A) Purification of MPrimPol$_{1\text{-}338}$. MPrimPol$_{1\text{-}338}$ was purified in four chromatography steps and the SDS-PAGE analysis of samples from fractions at each stage is shown (39.2 kDa). Unlike purification of the human polymerase domain, MPrimPol$_{1\text{-}338}$ does not exhibit any proteolytic breakdown products. **(B)** MPrimPol$_{1\text{-}338}$ was incubated with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. MPrimPol$_{1\text{-}338}$ is proficient at extending an undamaged oligonucleotide substrate using dNTPs. MPrimPol$_{1\text{-}338}$ was also analysed in the absence of dNTPs and no activity was observed.

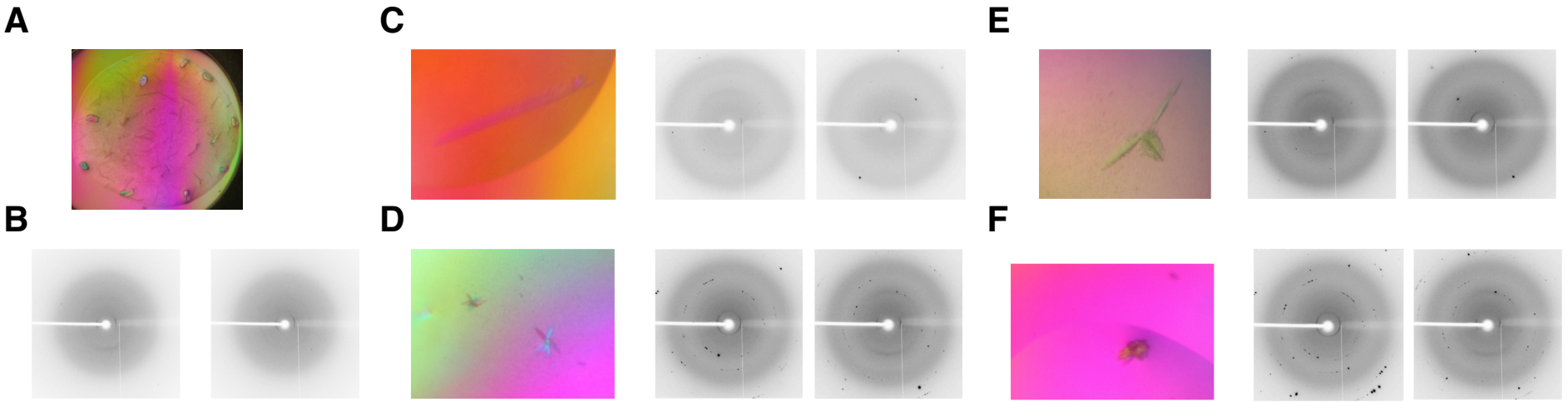


Figure 3.21. Crystallographic screening of MPrimPol$_{1\text{-}338}$.
(A) MPrimPol$_{1\text{-}338}$ was screened at 8.5mg ml^{-1} at a molar ratio of 1:1.5 with DNA in the presence of 4 mM $MgCl_2$. Reservoir Condition 0.2M $(NH_4)_2HPO_4$, 40% MPD. Crush test confirmed these crystals to be salt. **(B)** An example of an ice crystal formation test. Mother Liquor conditions were supplemented with 20% ethylene glycol as cryoprotectant and diffraction patterns are taken at ϕ angles of 0 and 90°**.** The diffuse pattern shown here with no apparent diffraction suggests that there are no ice crystals. **(C)** MPrimPol$_{1\text{-}338}$ was screened up at 7 mg ml^{-1} at a molar ratio of 1:1.5 with DNA in the presence of 4 mM $MgCl_2$. Crystals were present in the reservoir condition 0.09M NPS ($NaNO_3$; Na_2HPO_4; $(NH_4)_2SO_4$ mix), 0.1M Tris Bicine (pH 8.5), 30% PEG (PEG MME 500; PEG 20,000 mix). X-ray diffraction patterns were taken ϕ angles of 0 and 90° and the evident high intensity spots at mid-high resolution are indicative of salt crystals. **(D)** Mouse 1-338 PrimPol was screened up at 7 mg ml^{-1} at a molar ratio of 1:1.5 with DNA in the presence of 4 mM $MgCl_2$. Crystals were formed in the reservoir condition 0.09M NPS ($NaNO_3$; Na_2HPO_4; $(NH_4)_2SO_4$ mix), 0.1M Tris Bicine (pH 8.5), 30% precipitant (MPD; PEG 1000; PEG 3350 mix). X-ray diffraction patterns were taken ϕ angles of 0 and 90° and the multiple high intensity spot pattern is evident of salt crystals that may contain protein. **(E)** MPrimPol$_{1\text{-}338}$ was screened up at 7 mg ml^{-1} at a molar ratio of 1:3 with DNA in the presence of 4 mM $MgCl_2$. Crystals were formed in the same reservoir condition as in (C) and show a similar salt diffraction pattern. **(F)** Mouse 1-338 PrimPol was screened up at 7 mg ml^{-1} at a molar ratio of 1:1.5 with DNA in the presence of 4 mM $MgCl_2$ and 5 mM dideoxyadenosine triphosphate (ddATP). Crystals were formed in the reservoir condition 0.09M NPS ($NaNO_3$; Na_2HPO_4; $(NH_4)_2SO_4$ mix), 0.1M Tris Bicine (pH 8.5), 30% precipitant (Glycerol; PEG4000 mix). The X-ray diffraction patterns taken ϕ angles of 0 and 90° are similar to that of (E). The DNA substrate used in these crystal screens is substrate C1 from Table 2.6.

The protein was also screened at 7 mg mL^{-1}, with a DNA concentration such that the molar ratio was adjusted to 1:3. Crystals developed in one of the same condition as above but again the diffraction pattern is indicative of salt crystals. Finally, the protein was screened in the presence of all of DNA, $MgCl_2$ and a non-hydrolysable adenosine triphosphate (ATP) analogue, adenylyl imidodiphosphate (AMP-PNP). Again, crystals developed in a similar condition as above; however, these crystals were again salt crystals.

As with human PrimPol, first 24 amino acids were then truncated to produce a more compact AEP polymerase domain, $MPrimPol_{24-338}$. $MPrimPol_{24-338}$, like $MPrimPol_{1-338}$, purified without any evident degradation products (Figure 3.22). This protein was then screened for crystal formation at a concentration of 16 mg mL^{-1}, as determined by PCT. Whilst no crystals were evident during initial screening, there was a significant level of phase separation in conditions containing 2.2 M ammonium sulphate and 0.2 M of various potassium salts (Figure 3.23A). Different potassium salts were tested in ammonium sulphate at a range of temperatures (4°C, 12°C, 20°C and 37°C) and it was discovered that small crystals became evident in potassium acetate when incubated for 1.5 hours at 37°C, then transferred to room temperature overnight (Figure 3.23B). Upon altering the salt concentrations, well-defined crystals were also evident at a higher potassium acetate concentration of 0.3 M (Figure 3.23C). The intention was for these crystals to be analysed by X-ray diffraction, however when they were transferred into a mother liquor solution containing 20% ethylene glycol, the crystals dissolved. These crystals were under a ‘skin’ layer of protein that is often present in sitting drops with a high concentration of protein and I believe that there was a high local concentration of protein under this skin layer that stabilised the crystals; upon transferring to a mother liquor solution without protein, the crystals became destabilised and dissolved into the solution.

In addition to screening *apo* $MPrimPol_{24-338}$, crystal formation was also screened in the presence of DNA and 4 mM $MgCl_2$ in a 1:2 molar ratio. A single crystal developed in one of the conditions but X-ray diffraction data from this crystal was highly indicative of a salt crystal (Figure 3.23D).

As human hydrogen-deuterium exchange indicated that the first ~40 amino acids of PrimPol were not core amino acids and these residues align well with between human

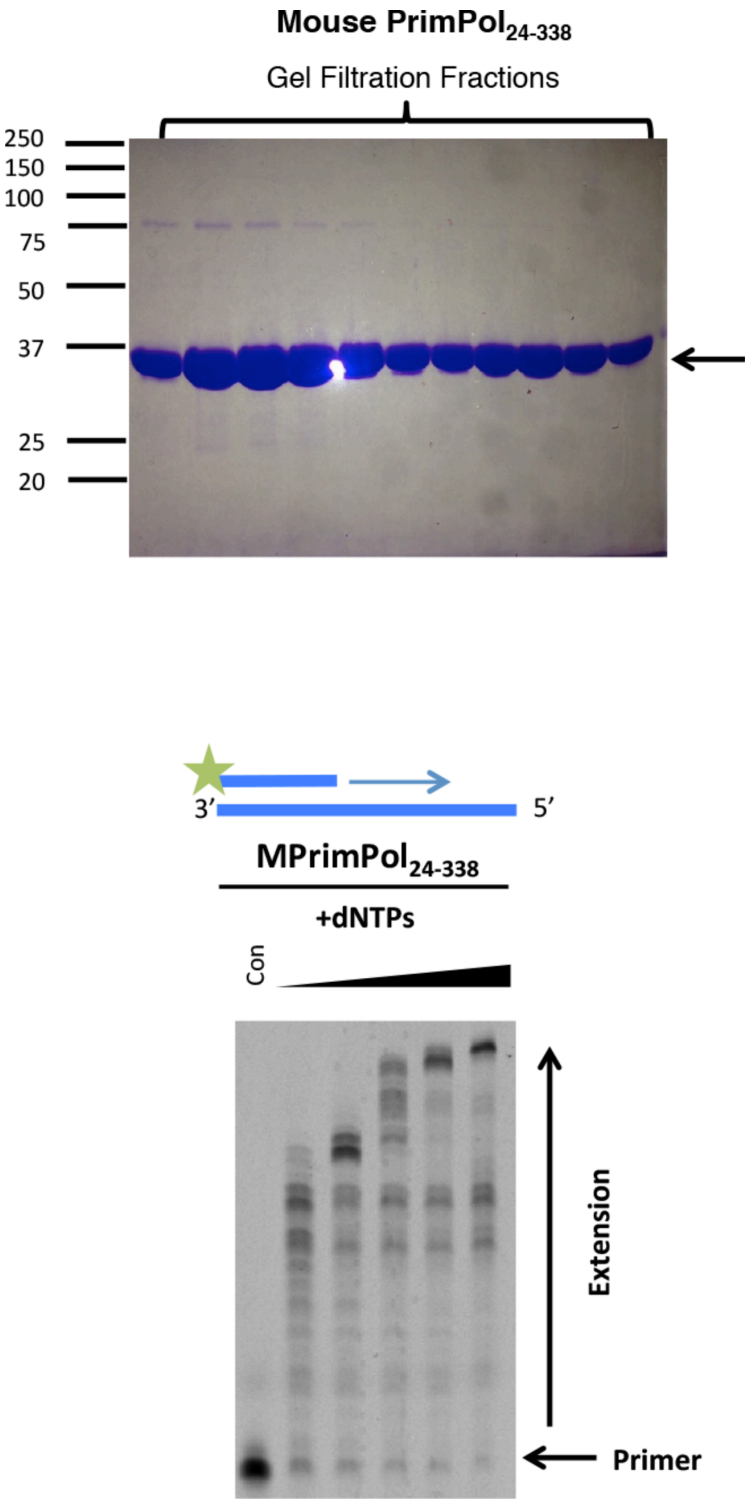


Figure 3.22. Purification and polymerase activity of MPrimPol$_{24-338}$.
(A) Purification of MPrimPol$_{24-338}$. MPrimPol$_{24-338}$ was purified in the same manner as MPrimPol$_{1-338}$ and the SDS-PAGE analysis of samples from fractions following size exclusion chromatography is shown (36.2 kDa). **(B)** MPrimPol$_{24-338}$ was incubated with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. MPrimPol$_{24-338}$ is proficient at extending an undamaged oligonucleotide substrate using dNTPs.

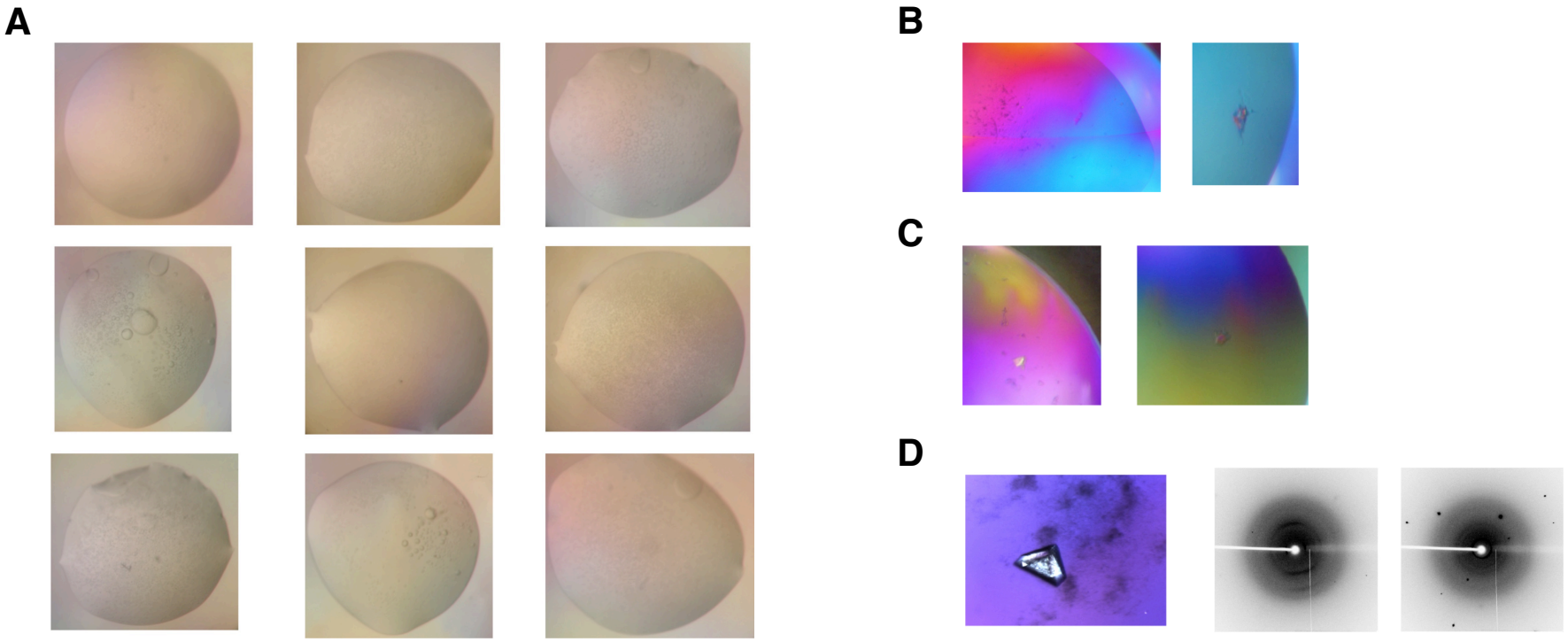


Figure 3.23. Crystallographic screening of MPrimPol$_{24-338}$.
(A) MPrimPol$_{24-338}$ screened at 16 mg/ml showed significant phase separation in the sitting drops with reservoirs containing 2.2 M ammonium sulphate, with a number of potassium salts. Potassium salts shown here from top left include 0.2 M KBr, 0.2 M KCl, 0.2 M K_3 Citrate, 0.2 M KH_2PO_4, 0.2 M KF, 0.2 M KCOOH, 0.2 M K_2HPO_4, 0.2 M KI, 0.2 M KNO_3, 0.2 M K/Na Tartrate. (B) Following screening of 2.2 M ammonium sulphate with different potassium salts at a number of temperatures, it was discovered that crystals could be grown in the sitting drops of a reservoir containing 2.2 M ammonium sulphate and 0.2 M potassium acetate when incubated at 37°C for 1 hour before incubation at 20°C. (C) Following variation of salt concentrations, crystals were also evident at 0.3 M potassium acetate. (D) MPrimPol$_{24-338}$ was screened up at 7 mg ml^{-1} at a molar ratio of 1:2 with DNA in the presence of 4 mM $MgCl_2$. Crystals were present in the reservoir condition 0.1 M Ammonium Acetate, 0.02 M Magnesium Chloride, 0.05 M Hepes-Na (pH 7), 5% (w/v) PEG 8000. X-ray diffraction patterns were taken ϕ angles of 0 and 90° and the evident high intensity spots at mid-high resolution are indicative of salt crystals. The DNA substrate used in these crystal screens is substrate C1 from Table 2.6.

and mouse PrimPol, we also removed the first 38 amino acids of Mouse PrimPol. This $MPrimPol_{39-338}$ was cloned and purified (Figure 3.24A). Upon testing for primer extension activity, the same unusual signature is apparent that was seen for human $PrimPol_{39-354}$ (Figure 3.24B). This suggests the removal of these 38 amino acids, in particular the amino acids between 24 and 38, has an effect on the structure of the AEP polymerase domain core. Screening this protein for crystals yielded no significant results.

3.13.1. Hydrogen-Deuterium Exchance Reveals the Unstructured Regions of the Mouse PrimPol AEP Polymerase Domain

In collaboration with the Skehel laboratory, hydrogen-deuterium exchange was also applied to the Mouse PrimPol AEP polymerase domain ($MPrimPol_{1-338}$). As with human PrimPol, as expected, the three highly conserved motifs, D114 and E116 of motif I, S167 and H169 of motif II, and D262 of motif III are all in low exchange regions of the protein (Figure 3.25). The exchange was high in the first 34 amino acids, but there was actually a protection from exchange from amino acid 35. This suggests that the removal of amino acids 35-38 may have an affect on the central, core structure of the protein, which would account for the unusual activity evident in Figures 3.19 and 3.24 in which the human and mouse PrimPol polymerase domains, respectively, are disrupted to some extent. Consequently, the disruption of the polymerase domain could lead to a more open, flexible conformation, which is even more disinclined to crystallise.

The hydrogen-deuterium exchange data also suggested that amino acids 314-323 are more exposed at the surface of the protein. As such, $MPrimPol_{1-314}$ was constructed to test whether the amino acids C-terminal of 314 are required for activity. This protein was expressed but was found to be mostly insoluble (Figure 3.26A). Although in the elution step of the affinity chromatography step during SDS-PAGE analysis there appears to not be any protein present, spectrophotometry confirmed the presence of a low level of protein. As such, this protein was concentrated and subjected to chromatography through the use of a Heparin column and the fractions are shown in Figure 3.26A. This protein was found to be active (Figure 3.26B), thus the amino acids C-terminal of 314 are not essential for activity. However, as the vast majority of this protein was insoluble, the purification yields were far too low for crystallography, the yields were less than 1 mg per litre of culture, compared to yields of over 50 mg per litre of culture for $MPrimPol_{1-338}$, $MPrimPol_{24-338}$ and $MPrimPol_{39-338}$. Whilst amino acids

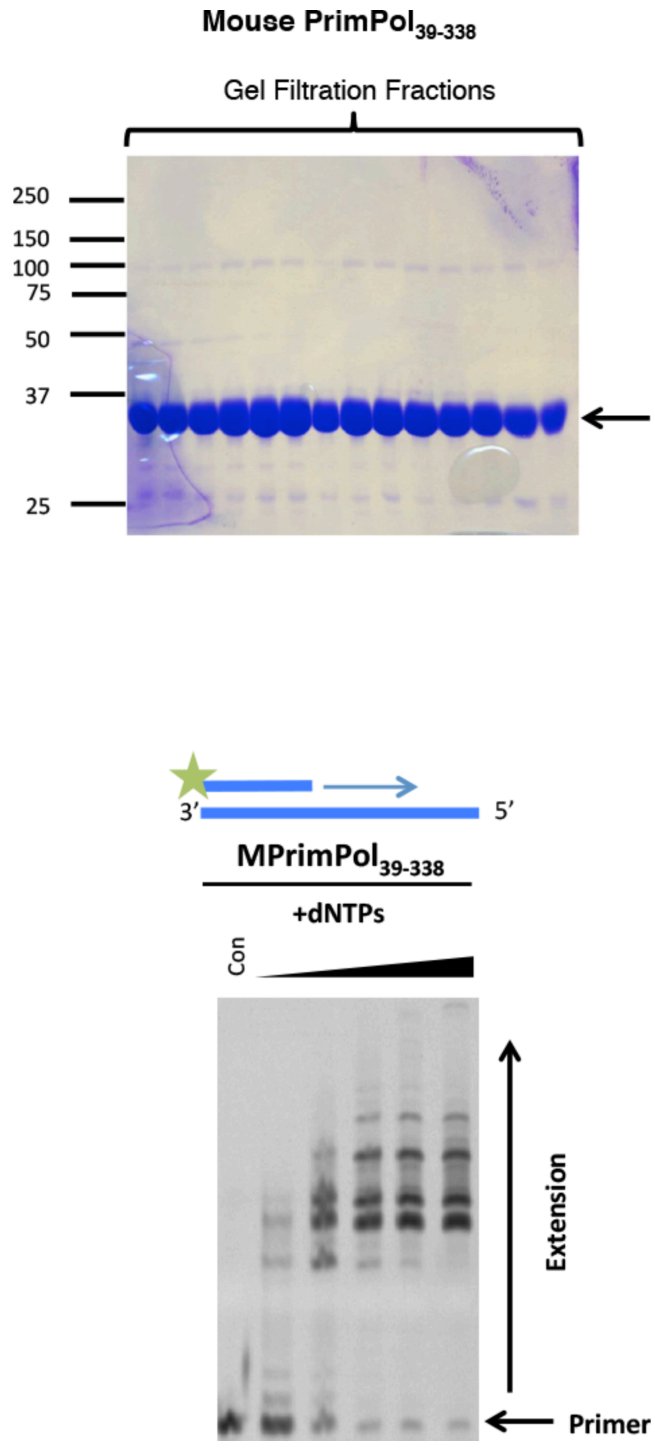


Figure 3.24. Purification and polymerase activity of MPrimPol$_{39-338}$.
(A) SDS-PAGE analysis of MPrimPol$_{39-338}$ showing samples from the protein-containing fractions from the size exclusion chromatography of MPrimPol$_{39-338}$ (34.5 kDa). **(B)** MPrimPol$_{39-338}$ was incubated with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points, along with a control containing no protein. MPrimPol$_{39-338}$ has a similar, strange signature of activity as PrimPol$_{39-354}$, stalling at various points when extending the substrate primer and having very little primer extended fully, even after 60 minutes.

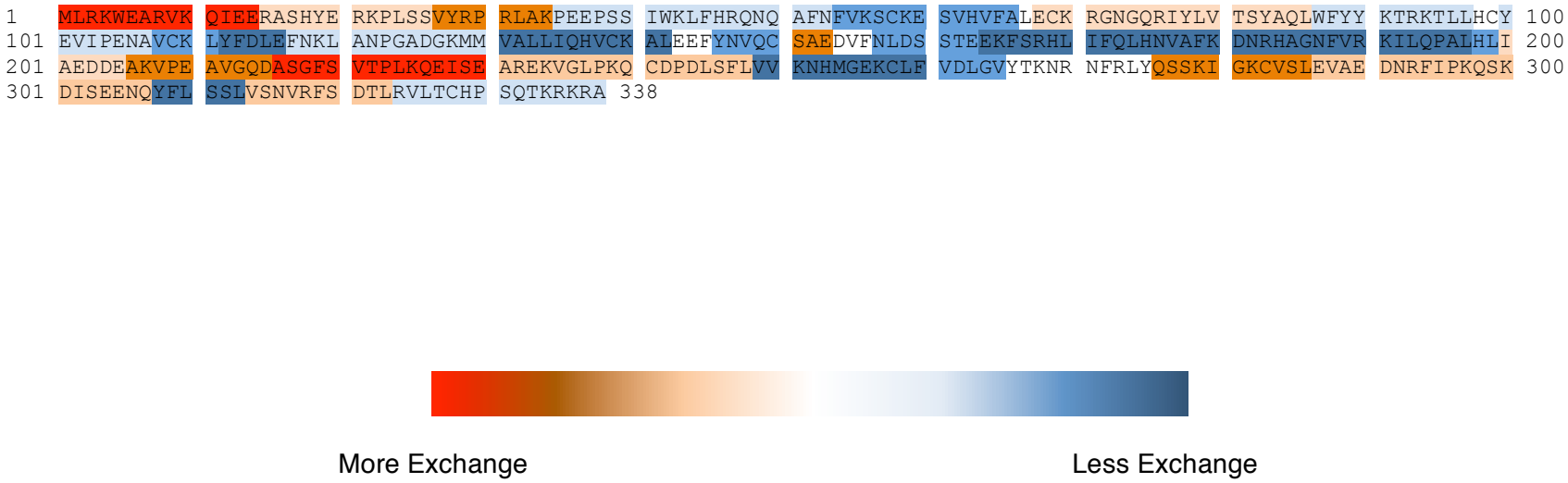


Figure 3.25. Hydrogen Deuterium Exchange of MPrimPol$_{24\text{-}338}$.
Protein areas of high isotopic exchange are indicated in red and those areas of lower isotopic exchange are indicated in blue. Low exchange is indicative of the more protected, central regions of protein, whilst high exchange is indicative of more exposed, external regions of protein. These experiments were carried out in the laboratory of J. Mark Skehel.

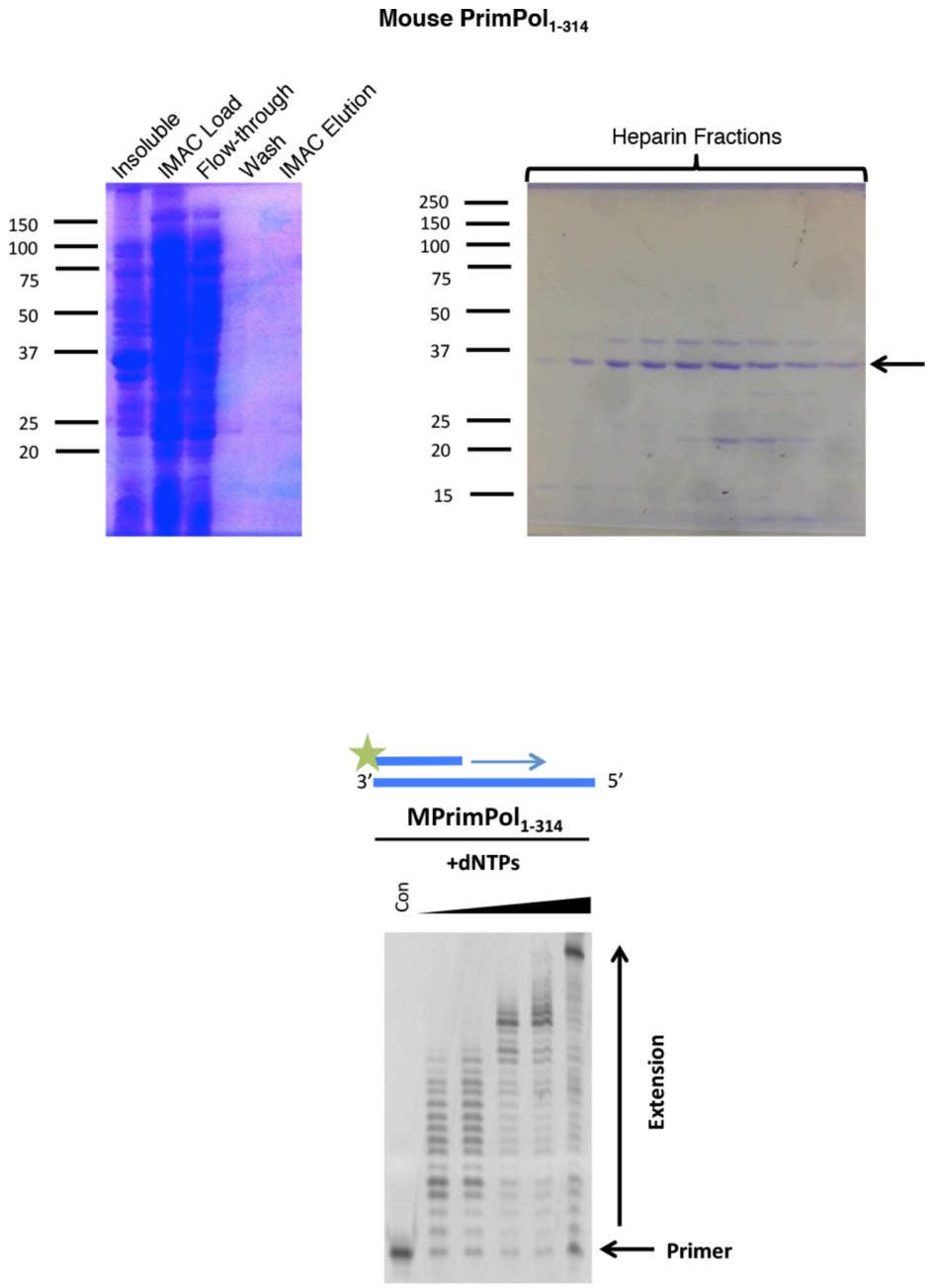


Figure 3.26. Purification and polymerase activity of MPrimPol$_{1\text{-}314}$.
(A) SDS-PAGE analysis of MPrimPol$_{1\text{-}314}$ showing samples from insoluble and soluble (load) fractions from the growth and expression of MPrimPol$_{1\text{-}314}$ showing MPrimPol$_{1\text{-}314}$ to be mostly insoluble (36.4 kDa). Although little protein is evident in the SDS-PAGE analysis of the elution of protein from the IMAC affinity chromatography column, spectrophotometry analysis indicated the presence of protein. Following concentration and purification by heparin column chromatography, some MPrimPol$_{1\text{-}314}$ protein was evident **(B)** MPrimPol$_{1\text{-}314}$ was incubated with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points, along with a control containing no protein. MPrimPol$_{1\text{-}314}$ is active, but present in too low a yield for crystallographic analysis.

315-338 may not be essential for catalysis, they are most likely involved in protecting the protein from aggregation and entering inclusion bodies in *E. coli*.

3.14. Summary and Discussion

Recent studies have identified that PrimPol is a novel eukaryotic AEP superfamily polymerase capable of performing both priming and primer extension synthesis (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Rudd et al., 2013; Wan et al., 2013). In this chapter, we have dissected the molecular activities associated with the major structural elements of human PrimPol and determined the cellular requirements for these elements *in vivo* for maintenance of replication fork speed. The DNA polymerase activity of PrimPol appears to be independent of the zinc finger domain. However, a catalytically active PrimPol lacking the zinc finger domain ($PrimPol_{1-354}$) shows enhanced relative polymerase compared to the wild-type enzyme, and in the next chapter this apparent polymerase activity regulation by the zinc finger is explored further.

An intact zinc finger is essential for PrimPol's primase activity *in vitro*, consistent with reports for a similar requirement for the highly related zinc finger domain of herpesvirus UL52 (Biswas and S. K. Weller, 1999; Y. Chen et al., 2005). Zinc-binding elements are found in most DNA primases (Section 1.7) and play critical functions in such primases (Kuchta and Stengel, 2010). It has been proposed that this structural element is required for template recognition and regulation of primer length synthesis (Corn et al., 2005). Interestingly we find that, out of the four homopolymers tested, PrimPol only has activity priming opposite a poly(dT) polymer. This specificity for homopyrimidine templates is not uncommon among AEPs. For example, the replicative AEPs from *M. jannaschii* and *T. kodakaraensis* are only able to prime on homopyrimidine templates and not homopurine templates and the primase from *P. furiosus* specifically primes on poly(dT) templates (Bocquier et al., 2001; Desogus et al., 1999; Galal et al., 2012b). The *S. solfataricus* preferentially primes in a synthetic thymine-containing bubble structure that mimics early replication intermediates (De Falco et al., 2004).

Removal of the last 73 amino acids of human PrimPol at the C-terminal region proximal to the zinc finger ($PrimPol_{1-487}$) did not abolish primase activity, which suggests that the C-terminus is not an absolute requirement for activity; however, activity was reduced to some extent suggesting that there likely are amino acids in this region responsible for

the structural integrity of the zinc finger. As they are not required for catalysis, we investigate the potential requirements of the final 80 amino acids in protein-protein interactions in Chapter 6.

Complementation of UL52-knockout viral cells with UL52 containing a mutation of conserved zinc-chelating residues was unable to support viral growth (Biswas and S. K. Weller, 1999). We find, however, that a zinc-finger mutant of PrimPol not only successfully complements PrimPol$^{-/-}$ avian cells but in the case of PrimPol$_{1\text{-}354}$, in which the zinc finger is removed entirely, replication fork speed is actually increased relative to the wild-type. This UL52 protein containing a mutation in its zinc finger motif had a 92% reduction in forked DNA binding activity and had lost all of its primase activity (Biswas and S. K. Weller, 1999). Whilst the disruption of the zinc finger in humans indeed had a profound effect on primase activity, it is unlikely that binding to forked DNA is hindered due to the observed increase in polymerase activity that is evident when the zinc finger of PrimPol has been removed. Additionally, we have shown here that the polymerase domain of PrimPol binds both ssDNA and dsDNA in the absence of the zinc finger in the high-nanomolar to low-micromolar range.

The activity most closely associated with the AEP superfamily, namely *de novo* primer synthesis, is considered to be somehow different from primer extension activities associated with canonical polymerases. However, as discussed in section 1.7, primases should be considered to be polymerases that have evolved to catalyse a poorly processive extension reaction involving the synthesis a dinucleotide. To do this, they have the capacity to bind a single nucleotide as the “primer”, which provides the 3’OH to initiate attack on a second template-bound nucleotide to form the first phosphodiester linkage. As the zinc finger binds single-stranded DNA and its absence negates primer synthesis, we propose that the zinc finger acts in concert with the polymerase domain to stabilise the binding and configuration of the template, polymerase and nucleotides long enough to allow the production of the initial dinucleotide “primer” (Figure 3.27). This product is subsequently extended in a more processive polymerisation mode. The function of the zinc finger is explored further in Chapter 4.

To gain a clearer understanding of how the polymerase domain carries out its catalytic activities, we attempted the rational design of a minimal PrimPol polymerase domain

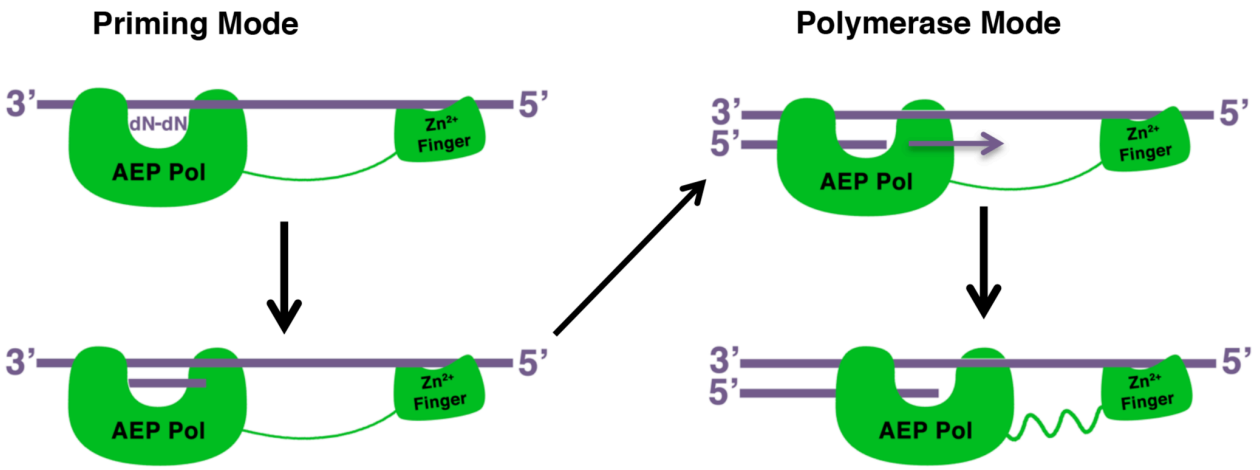


Figure 3.27. Model of the catalytic activities of PrimPol.
PrimPol has both DNA primase and polymerase activities. The zinc finger of PrimPol stabilises the binding of the polymerase domain on single-stranded DNA. The AEP polymerase domain coordinates two adjacent deoxynucleotides opposite the single-stranded DNA template and the stability afforded to this complex by zinc finger binding allows synthesis of a dinucleotide. This dinucleotide can then be extended in a more processive way to form longer products. PrimPol can also bind primer-template junctions to catalyse more canonical primer extension reactions. Both the polymerase and primase activities are intrinsic to the AEP domain. We postulate that the primase and polymerase activities of PrimPol are not discrete activities but rather represent different synthesis modes performed under different DNA binding conditions.

for human PrimPol to get an X-ray crystal structure of the protein. Hydrogen-deuterium exchange revealed potentially unstructured, flexible regions of the protein but we were unable to grow any adequate protein crystals. I presume this is in part a result of human PrimPol breaking down either during expression or purification. As such, we also attempted crystallography through the rational design of a minimal Mouse PrimPol polymerase domain. Although a number of crystals developed, these crystals were all determined to be salt crystals. A crystal structure would help determine whether a structural core is shared with the other AEP primases, given that PrimPol's zinc-chelating moiety is separate from the active polymerase domain. Other AEP primase structural cores consist of an ($\alpha_2\beta_2$) module and an RRM-like fold that is found in the palm region of other polymerases (L. M. Iyer et al., 2005). Given the high conservation of motifs I and III it would be rational to assume that PrimPol, like all other polymerases, employs the typical di-metal-ion nucleotidyl transferase reaction with the amino acids from motif II coordinating an incoming nucleotide. An *apo* crystal structure would help determine which amino acids are in close proximity to these conserved motifs in three-dimensional space and potentially support catalysis or template orientation. A crystal structure of PrimPol in the presence of DNA would illustrate how accommodating the active site is and how it is able to read through distortive DNA lesions such as (6-4)PPs (see Chapter 4).

CHAPTER 4

Characterisation of DNA Polymerase Activities of Human PrimPol

4.1. Introduction

PrimPol is present in both the nucleus and mitochondria of eukaryotic cells (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Rudd et al., 2013; Wan et al., 2013). The enzyme localises to nuclear chromatin during replication (Bianchi et al., 2013; Rudd et al., 2013; Wan et al., 2013) and this is more pronounced after treatment with damaging agents such as ultraviolet light (UV) or replication stalling drugs such as hydroxyurea (Bianchi et al., 2013). Deletion of PrimPol (PrimPol$^{-/-}$) induced replication fork slowing, which was much more pronounced when cells were UV irradiated (Bianchi et al., 2013). Knockout cells also exhibited increased formation of chromosomal breaks, particularly after aphidicolin treatment. UV irradiation can induce the covalent linkage of adjacent pyrimidines leading to the formation of CPDs and (6-4)PPs (Section 1.4.2). These helix-distorting lesions interfere with major biological processes, including DNA replication and transcription. PrimPol can perform translesion synthesis (TLS) bypass of the highly distorting (6-4)PPs but is also involved in replication through oxidative lesions, including 8-oxoguanine (8-oxo-G) (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Rudd et al., 2013). PrimPol is not epistatic with the Pol η-dependant CPD-bypass pathway and therefore appears to form an independent pathway required for the bypass of UV, and other lesions, during replication (Bianchi et al., 2013).

As the polymerase domains of TLS polymerases are more open and accommodating than high-fidelity, replicative polymerases to allow for replication past distortive DNA lesions (W. Yang, 2005), it is expected that that PrimPol has a similar active site. A more open, accommodating active site, however, can have a profound effect on a polymerase's fidelity, its tolerance of mismatched bases, and its kinetic parameters. In this chapter I investigate these properties of PrimPol and how they compare to other polymerases and what role the zinc finger plays in maintaining these properties.

Although PrimPol's polymerase activities appeared, initially, to be largely independent of the zinc finger domain (chapter 3), the data reported in this chapter suggests that the zinc finger domain plays important roles in the processivity and fidelity of DNA synthesis, as well as in other polymerase properties such as the tolerance of mismatched bases and the read-through of DNA damage. It is demonstrated here that a catalytically active fragment of PrimPol, containing only the AEP domain, catalyses TLS through both major UV-induced DNA lesions, CPDs and (6-4)PPs.

4.2. Processivity of PrimPol on DNA Templates

We analysed the number of nucleotides incorporated by human PrimPol in a single association-dissociation event to measure the processivity of this polymerase. In order to determine the processivity of PrimPol, we performed primer extension assays in the presence of an excess of DNA (sonicated herring sperm DNA), which acted as a "substrate trap" for polymerases that dissociate from the primer-template substrate (Washington et al., 1999). Using this approach, we observed that PrimPol inserts up to 4 nucleotides opposite an undamaged template, rarely inserting more than this (Figure 4.1A-B). It is worth noting that not all proteins present insert 4 nucleotides, with about equal efficiency in incorporating 1, 2 or 3 nucleotides. When the zinc finger domain was removed from PrimPol, the maximum number of nucleotides incorporated remained at ~4 but the average number of nucleotides incorporated was higher, with fewer polymerases inserting 1-3 nucleotides (Figure 4.1A-B). Processivity assays were carried out in the presence of Taq polymerase to ensure that it was not the trap DNA that kept the polymerase from inserting more than four nucleotides. We observe that Taq polymerase possesses high processivity and incorporates nucleotides up to the end of the template DNA (Figure 4.1C). Therefore, the zinc finger appears to play a role in regulating the number of nucleotides that PrimPol inserts, possibly by stabilising the interaction of the polymerase domain with the template.

4.3. Single Incorporation Kinetics of Wild-Type PrimPol

To quantify the processivity of PrimPol in terms of reaction rate, a single-incorporation kinetic analysis was performed to determine the catalytic efficiency (k_{pol}) and binding constants for dATP ($K_{D(dNTP)}$) for the incorporation of dATP opposite a templating thymine (see Section 2.3.6). This method is based on the quantification of nucleotide incorporation given a range of concentrations of dATP. We found the wild-type k_{pol} to be 6.98 min^{-1} and the $K_{D(dNTP)}$ to be 15.51 μM (Figure 4.2). PrimPol is an extremely slow enzyme, for comparison the k_{pol} of base excision repair enzyme DNA polymerase β is 1,944 min^{-1} (Nemec et al., 2014) and the k_{pol} of replicative enzyme pol ε is 16,500 min^{-1} (Zahurancik et al., 2014). The k_{pol} of Y-family polymerase pol ι is closer to the rate of PrimPol but is still quicker, at a rate of 37.8 min^{-1}(Washington et al., 2004). The $K_{D(dNTP)}$ for pols β, ε and ι were in the same range as that of PrimPol, 5.7 μM, 33 μM and 4.9 μM respectively (Nemec et al., 2014; Washington et al., 2004; Zahurancik et al., 2014). Representative examples of single incorporation experiments are shown in appendix III.

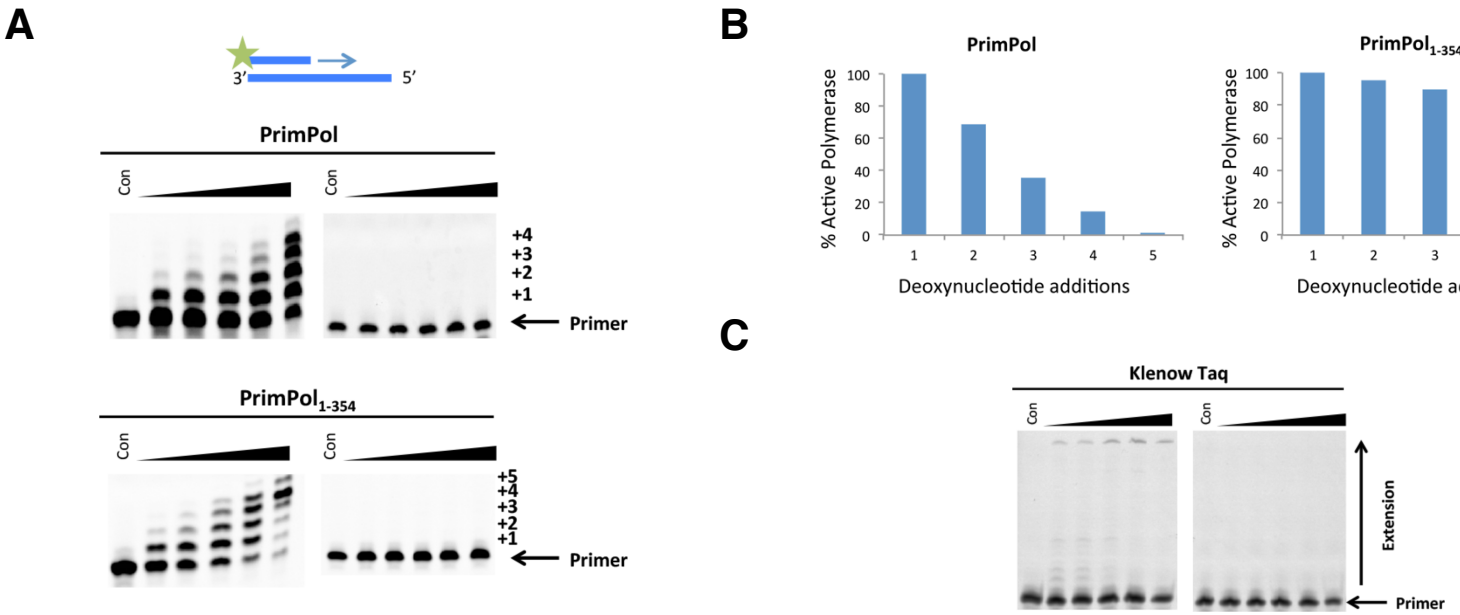
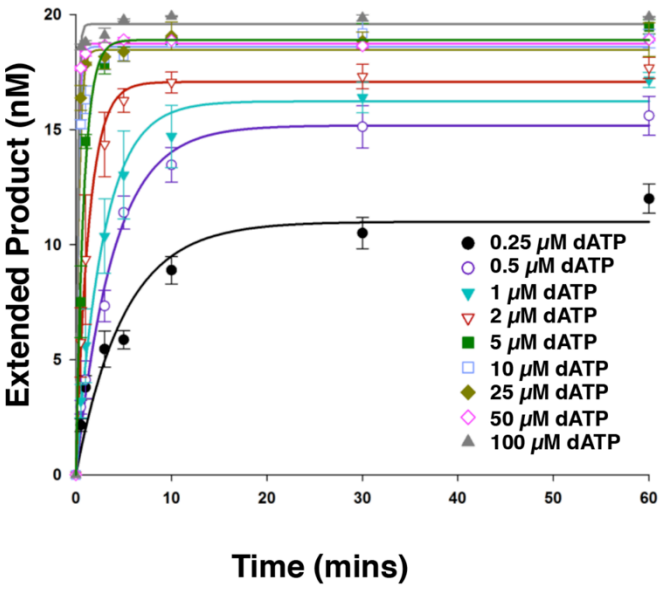


Figure 4.1. Processivity of Human PrimPol.
PrimPol was preincubated for 30 minutes at 37°C with an undamaged fluorescent DNA primer-template substrate (substrate 1 from Table 2.6) to allow PrimPol to bind to the DNA. The reaction was initiated through the addition of dNTPs and an excess of sonicated herring sperm DNA trap and time points taken at 15 s, 30 s, 60 s, 120 s and 360 s. **(A)** After 360 s, full length PrimPol incorporates up to 4 nucleotides opposite the template but there is a significant fraction of enzyme that incorporates only 1, 2 or 3 nucleotides. To confirm that the trap prevents polymerase extending from a second template, the trap was also added into the preincubation mix with PrimPol and the DNA substrate. This reaction was supplemented with dNTPs and there is no extension, thus successfully exhibiting the effectiveness of the trap. PrimPol$_{1\text{-}354}$ also predominantly incorporates up to 4 nucleotides but, unlike the full-length, there are fewer polymerases that only incorporate 1, 2 or 3 nucleotides. The effectiveness of the trap was also successfully confirmed for the 1-354 construct. **(B)** Percentage of PrimPol molecules incorporating at least n dNTPs for each of full-length PrimPol and PrimPol$_{1\text{-}354}$ was calculated using the equation in Section 2.3.4. **(C)** As a processive polymerase control, Klenow Taq polymerase was also assessed for its processivity. Klenow Taq extends DNA the full length of the template when preincubated with DNA. In the case of Klenow Taq, less substrate was extended because less Klenow Taq polymerase molecules (20 nM) was present than substrate (60 nM) and proteins being quenched when sonicated herring sperm DNA. The effectiveness of the trap in the case of Klenow Taq polymerase was also confirmed by the preincubation of Klenow Taq with herring sperm DNA prior to addition of dNTPs.

A



B

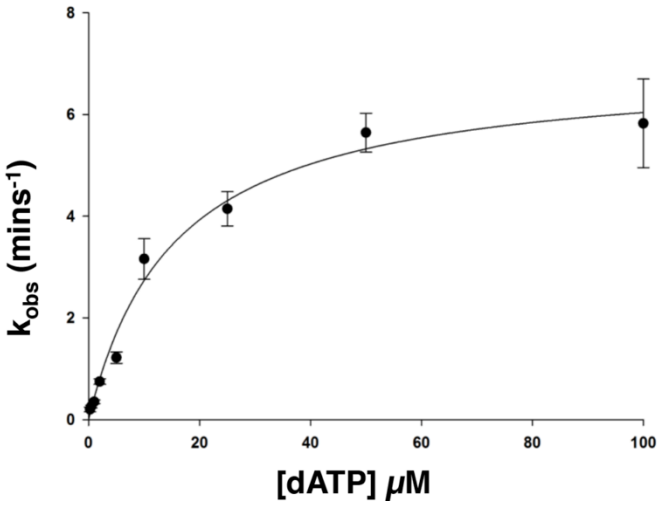


Figure 4.2. Single incorporation kinetics of PrimPol.
Single turnover kinetics of wild-type PrimPol were used to determine catalytic efficiency (k_{pol}) and dNTP binding efficiency ($K_{D(dNTP)}$). **(A)** The concentration of extended DNA product, as determined by electrophoresis, was plotted against time and fit to an exponential curve for a number of dATP concentrations. These data were fit to exponential curves as described in the equation in Section 2.3.6. and k_{obs} was determined. Experiments were carried out in triplicate and error bars represent standard deviations. **(B)** k_{obs} was subsequently plot against dATP concentration for PrimPol and these data were fit to hyperbolic curves as described in Section 2.3.6. to determine k_{pol} and $K_{D(dNTP)}$. k_{pol} for wild-type PrimPol was found to be 6.98 min^{-1} ± 0.40 and $K_{D(dNTP)}$ was 15.51 μM ± 2.71.

4.4. The Fidelity of PrimPol

Unlike the replicative polymerases, PrimPol lacks a 3'-5' exonuclease activity and as PrimPol is a poorly processive enzyme similar to other TLS polymerases, which themselves have low fidelity, we decided to measure the fidelity of PrimPol as it is potentially an error-prone DNA polymerase. We measured fidelity through a series of nucleotide incorporation experiments opposite oligonucleotides with different templating bases. For each templating nucleotide, adenine, cytosine, guanine or thymine, PrimPol's ability to incorporate an incoming dATP, dCTP, dGTP or dTTP was assessed. Wild-type PrimPol in these experiments exhibits little aberrant incorporation of nucleotides (Figure 4.3A). There is some evidence of incorporation of a third cytosine nucleotide at the N+3 position when incorporating cytosine opposite two templating guanines followed by a cytosine, suggesting therefore that PrimPol could incorporate cytosine opposite another cytosine if it was already extending a template or, perhaps more likely, that the template DNA is 'scrunched' by PrimPol and one of the guanines is read twice as a templating base. $PrimPol_{1-354}$ that lacks the entire zinc finger domain unexpectedly shows a significantly reduced ability to accurately select the incoming nucleotide (Figure 4.3B). $PrimPol_{1-354}$ also exhibits this incorporation of a third, incorrect nucleotide when extending a primer and it is apparent in the case of incorporation of guanine, cytosine and adenosine. Also, this polymerase domain of PrimPol is evidently capable of insertion of an adenine nucleotide opposite cytosine. This observation suggests the potential importance of the zinc finger domain in influencing the fidelity of PrimPol's polymerase activity.

This initial examination of the fidelity of PrimPol is on a limited number of templates in a manner that was not quantified. Subsequent experiments by Dr. Stanislaw Jozwiakowski from our laboratory have offered additional insights into the fidelity of PrimPol. The measurement of fidelity using a plasmid-based gap-filling assay using plasmid pSJ4-lacZa demonstrates the error rate of human PrimPol to be 1.0×10^{-4}, this is almost an order of magnitude more frequent than the 3.6×10^{-5} error rate of the exo^- Klenow fragment of Taq polymerase from the same method (Guilliam et al., 2015), but it remains less error prone than Y-family TLS polymerases pol η, κ and ι (Matsuda et al., 2001; Ohashi et al., 2000; Washington et al., 2004), and is in the same range as Dpo4 (Fiala and Suo, 2004). Sequencing of errors introduced by PrimPol revealed that insertions and deletions, so called indel mutations, accounted for ~65% of these

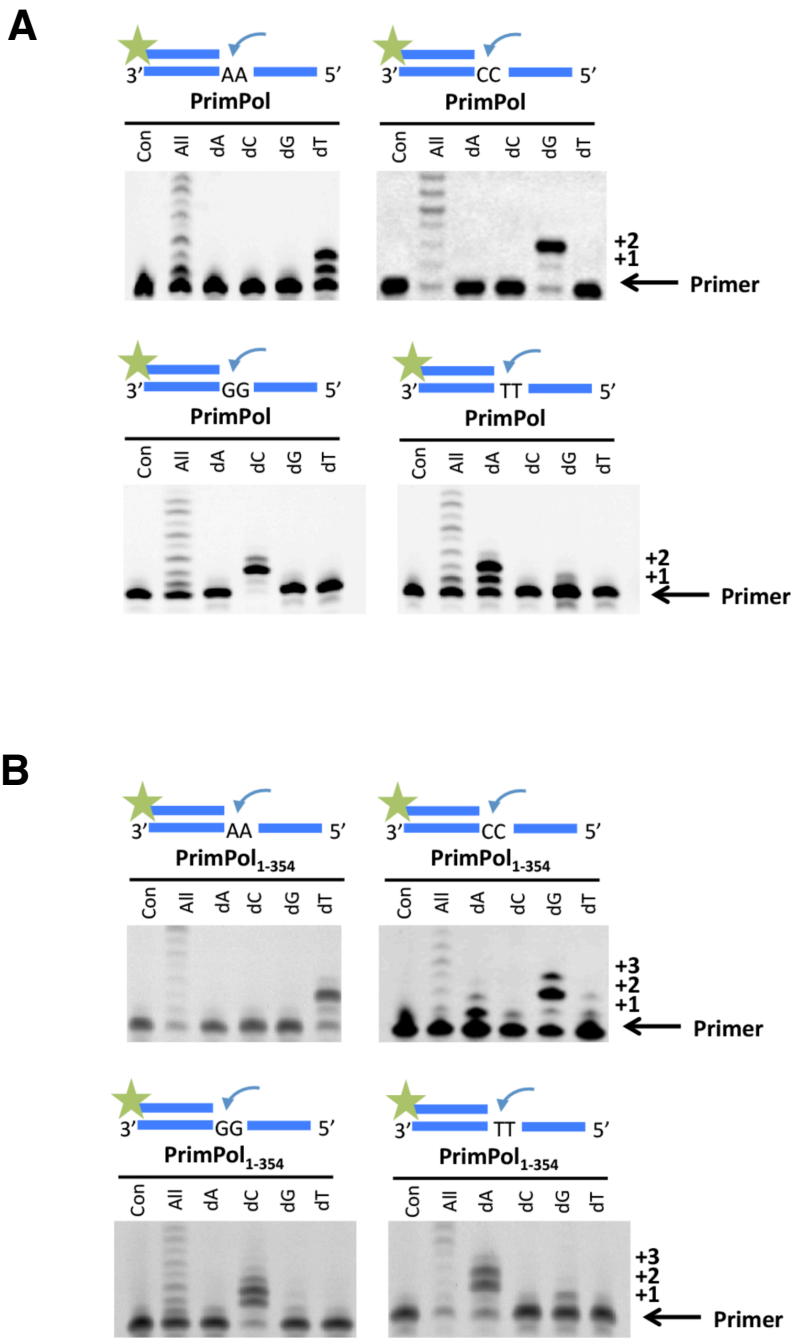


Figure 4.3. Single incorporation fidelity of PrimPol.
(A) PrimPol was incubated for 5 minutes with individual dNTPs as indicated, all four dNTPs or a no dNTP control along with substrates 9-12 from Table 2.6 to allow incorporation of single nucleotides opposite two templating bases for each individual base. PrimPol exhibits standard Watson-Crick base pairing opposite each template nucleotide but appears to erroneously incorporate nucleotides at position N+3 when incorporating cytosine opposite two guanines, with the third templating base in this case being cytosine. **(B)** $PrimPol_{1\text{-}354}$ was also tested for its fidelity in the same manner as PrimPol. $PrimPol_{1\text{-}354}$ exhibited more misincorporations relative to the wild-type protein; $PrimPol_{1\text{-}354}$ incorporates adenine opposite a template cytosine with relatively high proficiency. $PrimPol_{1\text{-}354}$ also exhibits more pronounced N+3 misincorporations, utilising each of guanine, cytosine and adenine opposite an incorrect N+3 templating base.

mutations (Guilliam et al., 2015). By far the most significant base substitution mutation was an G→A/C→T mutation, accounting for 25% of total mutations caused by PrimPol, this is the same type of mutation we see exhibited by $PrimPol_{1-354}$ (Guilliam et al., 2015)(Figure 4.3).

4.5. Base Mismatch Tolerance of PrimPol

Taking into consideration that PrimPol can occasionally incorporate the incorrect nucleotides opposite a templating base, we went on to measure whether PrimPol could extend from a range of templates that had a mismatched base at its primer-template junction, similar to the products you would create following a mismatched base incorporation. PrimPol was unable to extend from a range of mismatched bases (Figure 4.4); there was only evidence of very weak extension from an A/A mismatch, an A/G mismatch and a T/G mismatch. In contrast, the polymerase domain of PrimPol, $PrimPol_{1-354}$, exhibited the ability to incorporate bases following a mismatch opposite a range of primer-templates (Figure 4.5; summarised in Table 4.1).

Many of these incorporations are Watson-Crick base pairs with respect to the next templating base, suggesting that the polymerase domain of PrimPol is accommodating this mismatch and extending it in a template-dependent manner. However, there are a few cases in which guanine is incorporated where cytosine is not the next templating nucleotide but rather the one in position N+2, suggesting that the polymerase domain is scrunching the template and incorporating guanine opposite the downstream cytosine. *In vivo* this would result in deletions. Notably, deletion mutations accounted for 44% of total mutations caused by PrimPol in the gap-filling plasmid experiments carried out in our laboratory (Guilliam et al., 2015).

4.6. Template-independent Primer Extension Activity in the Presence of Manganese

The flexible extension activities reported previously for NHEJ AEP polymerases includes limited template-independent terminal transferase activity on both ssDNA and blunt-ended dsDNA (Della et al., 2004; Pitcher et al., 2007). Such activities are often stimulated by the presence of manganese ions, which can accelerate the rate of DNA synthesis and, as a result, lower the overall fidelity of this process. We therefore tested if human PrimPol also possesses a similar terminal transferase activity in the presence

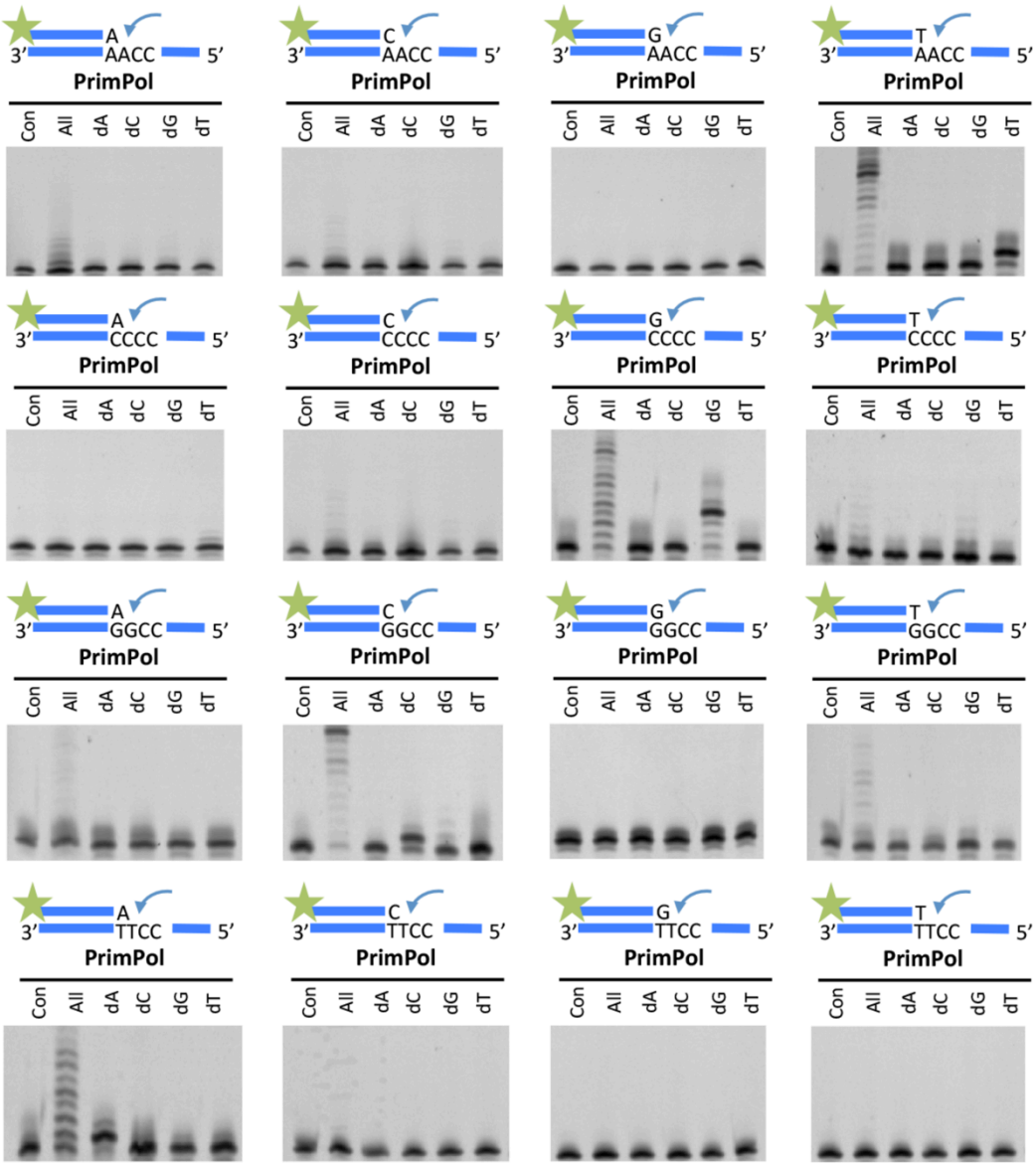


Figure 4.4. Base mismatch tolerance of wild-type PrimPol.
PrimPol was incubated for 5 minutes with individual dNTPs as indicated, all four dNTPs or a no dNTP control along with substrates 13-28 from Table 2.6. These substrates represent primers annealed to templates with the primer template junction representing each of the 16 possible combinations of base pairs, including the 4 Watson-Crick base pairs. PrimPol only extends the Watson-Crick base paired primer termini and will not extend from any base mismatches.

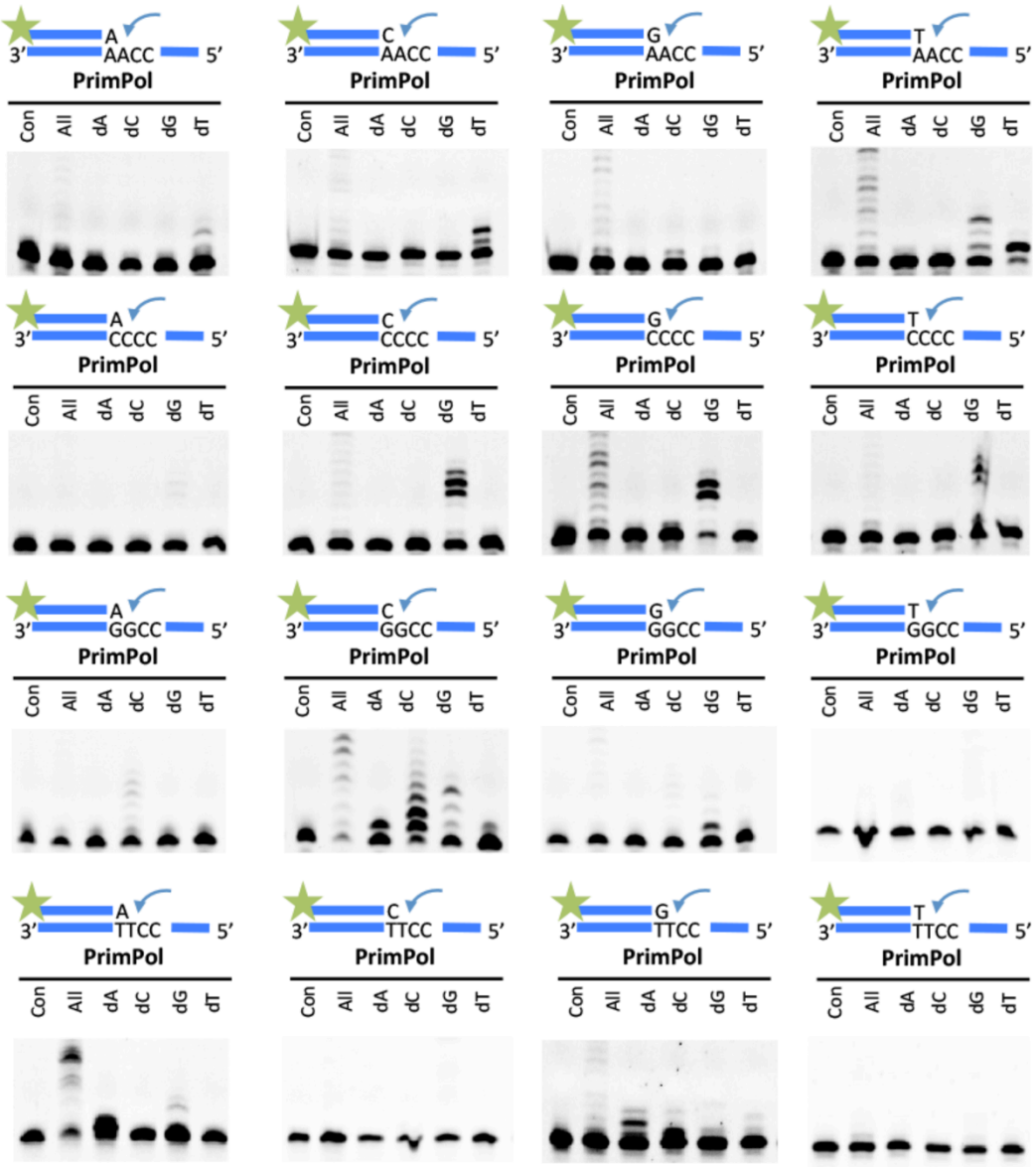


Figure 4.5. Base mismatch tolerance of PrimPol$_{1\text{-}354}$.
PrimPol$_{1\text{-}354}$ was analysed for its tolerance of base mismatches in the same manner as wild-type PrimPol in Figure 4.4. In addition to the standard Watson-Crick base pairs, PrimPol$_{1\text{-}354}$ shows tolerance for a number of other mismatched primer termini. The mismatch tolerance of PrimPol$_{1\text{-}354}$ is summarised in Table 4.1.

Sequence	Pairing Conformation	PrimPol		$PrimPol_{1-354}$	
		Tolerance	Nucleotide Incorporation	Tolerance	Nucleotide incorporation
3'---AACC----5' 5'---A	Frayed	-		+	T
3'---AACC----5' 5'---C	Disordered	-		+	T
3'---AACC----5' 5'---G	Frayed	-		+	C
3'---AACC----5' 5'---T	Watson-Crick	+++	T	+++	T>>G
3'---CCCC----5' 5'---A	Disordered	-		-	
3'---CCCC----5' 5'---C	Frayed	-		+	G
3'---CCCC----5' 5'---G	Watson-Crick	+++	G	+++	G
3'---CCCC----5' 5'---T	Disordered	-		+	G
3'---GGCC----5' 5'---A	anti-anti	-		-	
3'---GGCC----5' 5'---C	Watson-Crick	+++	C	+++	C>>A>G>>T
3'---GGCC----5' 5'---G	syn-anti	-		+	G
3'---GGCC----5' 5'---T	Wobble	-		-	
3'---TTCC----5' 5'---A	Watson-Crick	+++	A	+++	A>>>G
3'---TTCC----5' 5'---C	open	-		-	
3'---TTCC----5' 5'---G	wobble	-		+	A>C
3'---TTCC----5' 5’---T	wobble	-		-	

Table 4.1. Summary of the base mismatch tolerance of PrimPol and $PrimPol_{1-354}$. The conformations of each base pairing combination is described as resolved structurally (Johnson & Beese, 2004). Whilst PrimPol appears relatively intolerant to the extension of base mismatches, $PrimPol_{1-354}$ shows some tolerance to many of the mismatched base pairs.

of manganese. These experiments were carried out in collaboration with Dr. Stanislaw Jozwiakowski. Wild type, as well as all three variants, of human PrimPol showed no detectable terminal transferase activity when tested on a blunt ended ds DNA substrate (Figure 4.6). However, PrimPol's DNA synthesis activity appeared to be significantly increased when primer extension assays were performed in the presence of dNTPs (all) and manganese (Figure 4.6). To investigate if this increased activity resulted from altered fidelity, we repeated the assays with each of the dNTPs and observed that PrimPol catalyses multiple extension of the primer even in presence of a single nucleotide pool. This promiscuous DNA synthesis was most efficient when enzyme was utilising dATP or dTTP. In both cases, extension of primers by up to ~20 nucleotides was observed. Notably, this phenomenon was not observed when magnesium was used instead of manganese (result not shown). When a single incorporation of dCTP was investigated, we observed significantly slower extension rates but the enzyme was able to extend the primer to produce predicted full-length product. In contrast, incorporation of correct incoming nucleotide (dGTP) resulted in least pronounced primer extensions up to 16 nucleotides.

When this experimental strategy was repeated to study this promiscuous DNA polymerase activity in the $PrimPol_{ZF\text{-}KO}$ and $PrimPol_{1\text{-}487}$ variants, we observed minimal misincorporation of all four tested single nucleotides. Interestingly, when catalytically active fragment of $PrimPol_{1\text{-}354}$ was tested, we observed more efficient misincorporation compared to $PrimPol_{ZF\text{-}KO}$ and $PrimPol_{1\text{-}487}$ variants. However, $PrimPol_{1\text{-}354}$ fragment was able to misincorporate between 1 to 2 incorrect incoming nucleotides. We are unable to conclude if this unique enzymatic activity of human PrimPol is significant for biological function of this enzyme *in vivo*. We speculate that under the experimental conditions tested, the zinc finger module allowed manganese-dependent primer extension resulting in the synthesis that is non-complementary to the template strand. This hypothetical scenario is only possible when the enzyme was able to bind to a single-stranded template strand of the DNA primer-template substrate. This interesting observation has raised questions regarding the mechanism by which PrimPol binds to and coordinates the primer-template substrate during DNA polymerisation.

4.7. Translesion DNA Synthesis Activity of PrimPol

PrimPol has the capacity to bypass a number of replication stalling lesions, including UV and oxidative lesions (Bianchi et al., 2013; Garcia-Gomez et al., 2013). To address

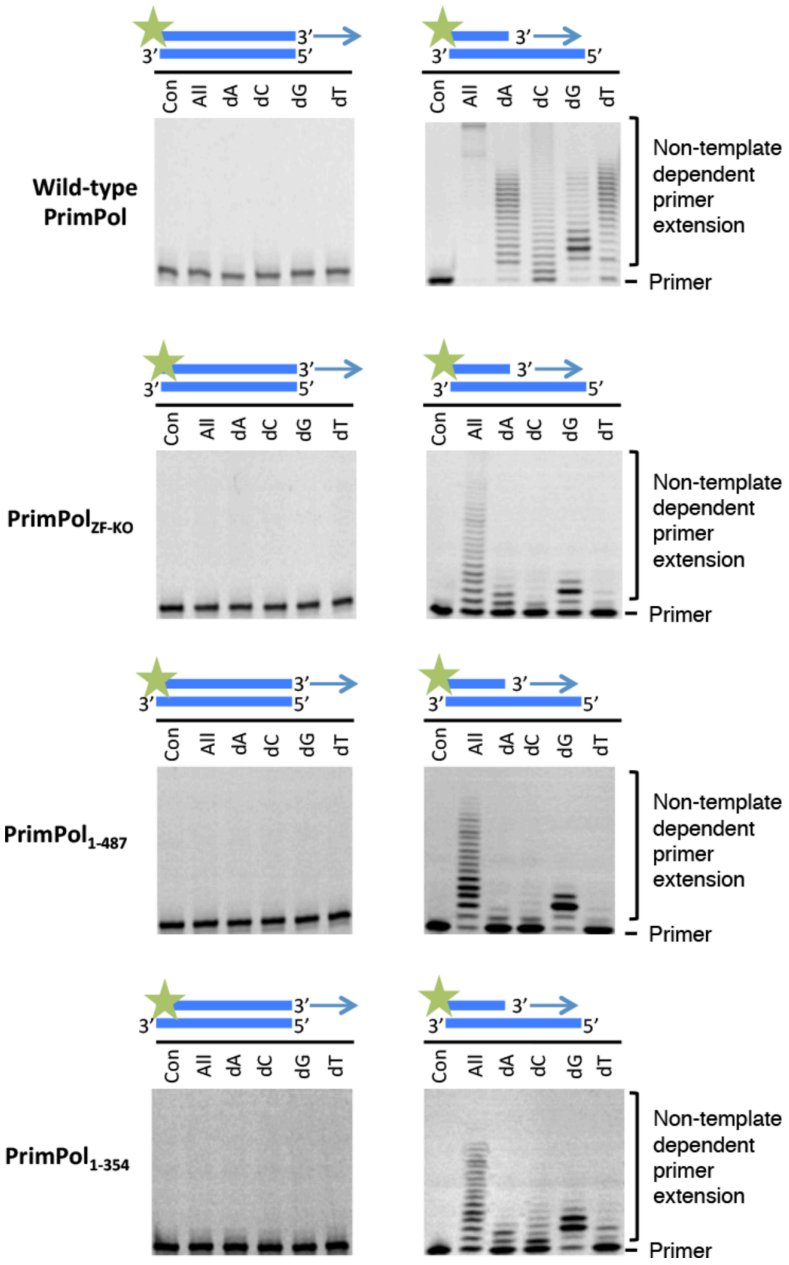


Figure 4.6. Template-independent extension by PrimPol in the presence of manganese.
Human PrimPol was incubated for 30 min with DNA substrate and each of the dNTPs in the presence of manganese. Wild-type PrimPol was unable to extend from a dsDNA template with a blunt end (substrate 3 from Table 2.6), but could extend from a primer annealed to an overhanging template (substrate 1 from Table 2.6), even synthesising long tracts of homopolymers. PrimPol$_{ZF-KO}$ could extend from an overhanging DNA template in the presence of manganese and, con- sistent with the wild-type, did not extend from a ds DNA substrate. The incorporation of 1 or 2 nucleotides of guanine or adenine opposite an over- hanging template suggests that PrimPol$_{ZF-KO}$ incorporates in a low-fidelity template-dependent manner when incubated with overhanging DNA. PrimPol$_{1–487}$ exhibited a highly similar terminal transferase activity spec- trum to the PrimPol$_{ZF-KO}$. In the presence of manganese, it incorporated bases opposite an overhang in a low fidelity, template-dependent manner. PrimPol$_{1–354}$ also exhibited low fidelity extension of a primer annealed to an overhanging template in the presence of manganese.

whether the TLS activities of PrimPol are independent of the zinc finger domain, we assayed the TLS activities of $PrimPol_{1-354}$ on a range of templates containing a variety of DNA modifications.

4.7.1. Error-Free Bypass of a CPD Photoproduct by PrimPol

As full-length PrimPol appears unable to read through CPDs *in vitro*, we assayed $PrimPol_{1-354}$ for TLS bypass of this common photolesion by primer extension of a primer annealed to a template containing a CPD. Notably, although $PrimPol_{1-354}$ slowed as it reached a CPD lesion, it showed the capacity to read-through this bulky lesion (Figure 4.7A). In contrast, wild-type PrimPol cannot perform TLS opposite this lesion in the presence of magnesium (Bianchi et al., 2013; Rudd et al., 2013) nor can replicative polymerases such as the Klenow fragment of Taq polymerase (Figure 4.7A). Single nucleotide incorporation assays were employed to determine which bases are incorporated opposite CPDs, revealing that it performs error-free bypass of this photolesion (Figure 4.7B). Similar TLS activity was also confirmed using $XPrimPol_{1-334}$, which was also capable of reading through CPDs in an error-free manner (Figure 4.7C). This TLS activity is similar to Pol η, a specialised TLS polymerase known to bypass CPDs by in an error-free manner (Biertümpfel et al., 2010; Masutani et al., 1999). As this read-through of a CPD is only evident in the absence of the zinc finger domain, it may be the case that PrimPol requires a conformational change or partner to assist in promoting this bypass activity *in vivo*.

4.7.2. Error-Prone Bypass of a (6-4) photoproduct by PrimPol

(6-4)PPs are a much more highly distorting UV lesion than CPDs (Section 1.4.2.). Pol η can incorporate bases opposite (6-4)PPs but incorporates an incorrect dG opposite the 3' dT of the lesion and is unable to subsequently extend from this product (R. E. Johnson et al., 2001). However, it has previously been reported that PrimPol exhibits both the ability to incorporate opposite and extend through (6-4)PPs (Bianchi et al., 2013). Although PrimPol's TLS activity was retained when the zinc finger was deleted (Figure 4.8A), the nucleotide incorporation signature was altered (Figure 4.8B), again consistent with the proposed regulatory role of the zinc finger domain. Wild-type PrimPol incorrectly incorporates thymine opposite the first dT of the (6-4)PP, whereas $PrimPol_{1-354}$ incorporates dA or dC opposite this base. If dA nucleotide is incorporated opposite the first thymine of the (6-4)PP, $PrimPol_{1-354}$ then incorporates either a dC or dT. If the first nucleotide is dC, it will then incorporate dA, dC or dT. To confirm whether

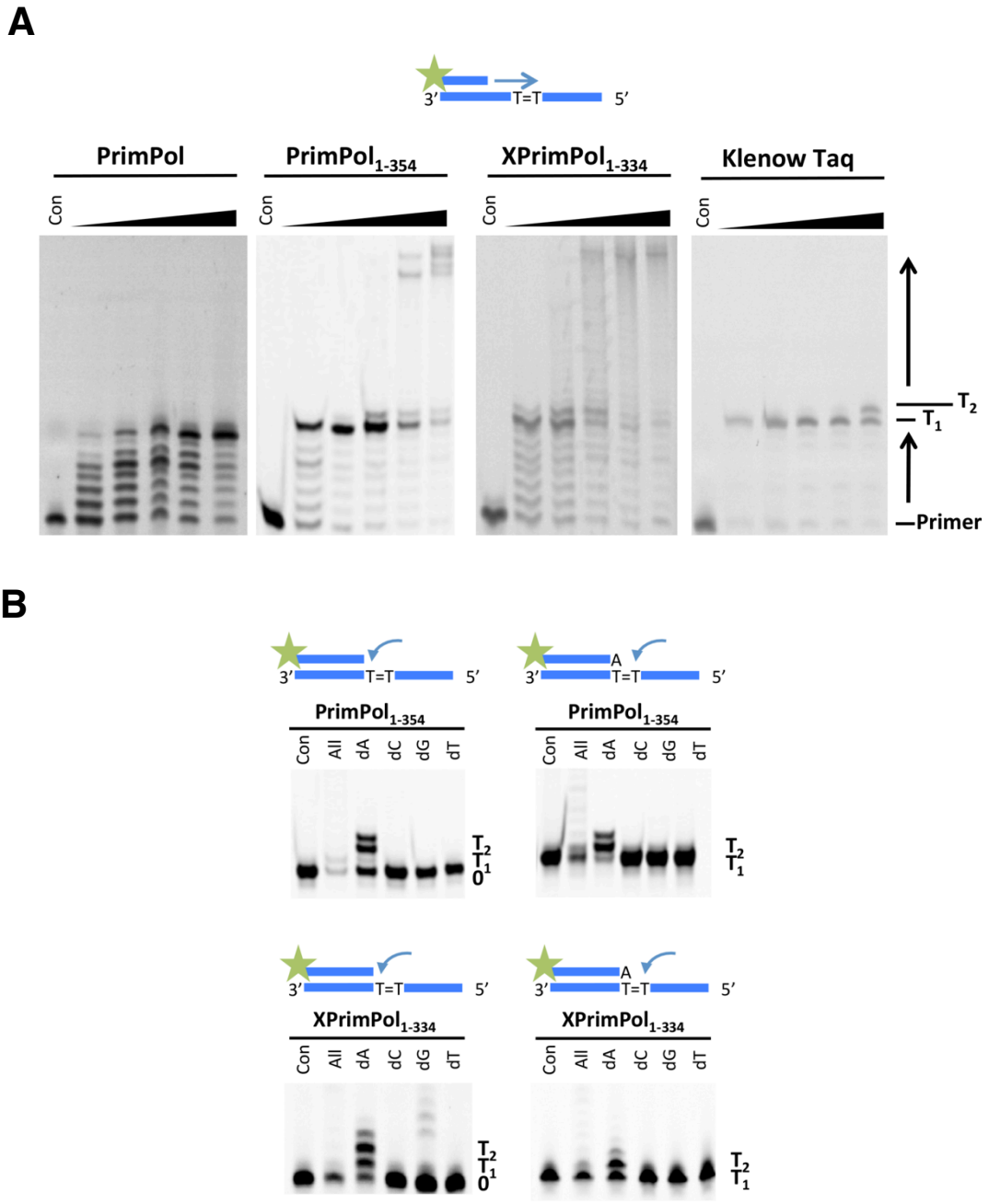


Figure 4.7. The polymerase domain of PrimPol can bypass a CPD lesion with high fidelity.
(A) PrimPol, PrimPol$_{1-354}$ and XPrimPol$_{1-334}$, as well as the replicative Klenow Taq polymerase, were incubated with dNTPs and a substrate with a CPD in its template (substrate 29 from Table 2.6) at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. As previously shown, PrimPol is unable to read-through a CPD. However the polymerase domains of both Human and X. tropicalis PrimPol, without the C-terminal zinc finger, are both capable of inserting nucleotides opposite, and extending from, a CPD lesion in vitro. Klenow Taq, a replicative polymerase, is incapable of reading through the distorting lesion. **(B)** PrimPol$_{1-354}$ and XPrimPol$_{1-334}$ were analysed for their fidelity opposite a CPD and were both shown to incorporate adenine nucleotides opposite both the 3'- and 5'-thymines of the CPD. Substrates 30 & 31 from Table 2.6 were used in this fidelity evaluation.

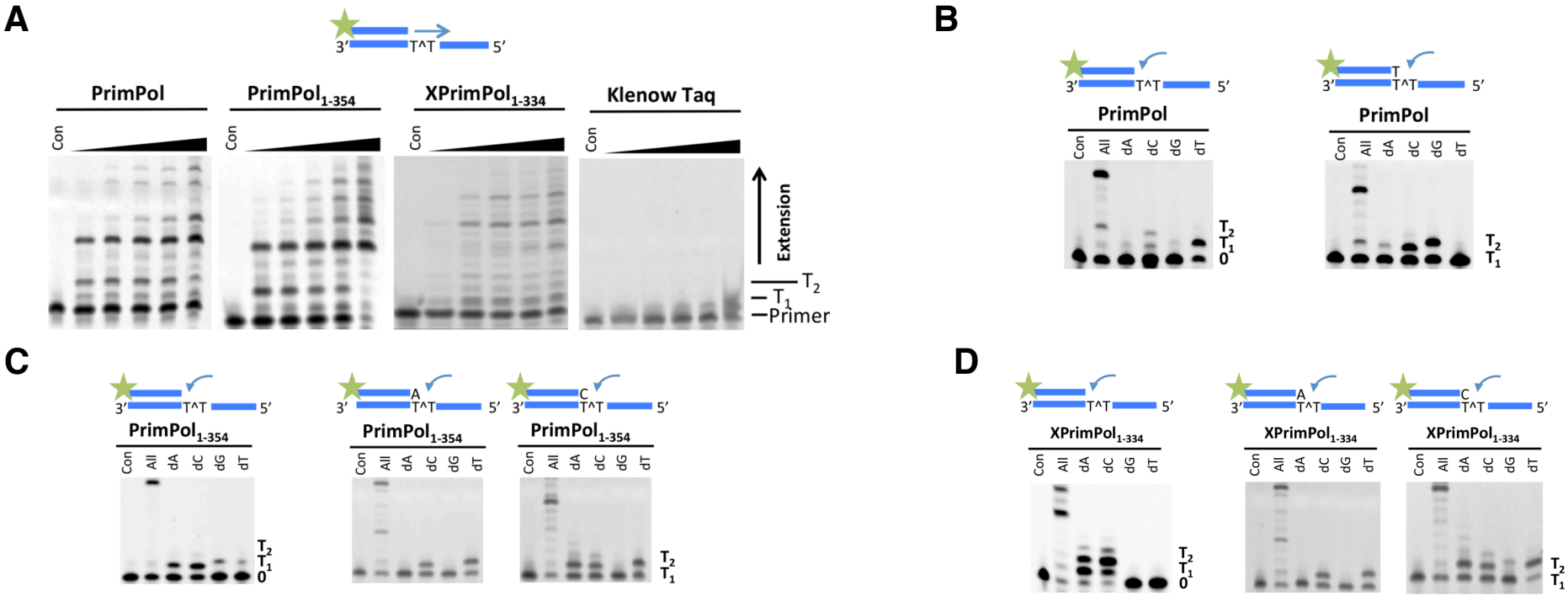


Figure 4.8. PrimPol can bypass a (6-4) photoproduct with low fidelity.
(A) PrimPol, PrimPol$_{1\text{-}354}$ and XPrimPol$_{1\text{-}334}$, as well as the replicative Klenow Taq polymerase, were incubated with dNTPs and a substrate with a (6-4)PP immediately downstream of the primer-template junction (substrate 33 from Table 2.6) at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. Both wild-type PrimPol and the polymerase domain of PrimPol are capable of proficiently reading through the (6-4)PP. Klenow Taq polymerase, however, is incapable of reading through this highly distortive lesion. (B) The fidelity of PrimPol when incorporating opposite a (6-4)PP was assessed. PrimPol predominantly incorporates a thymine nucleotide opposite the 3'-thymine of the (6-4)PP and then incorporates a cytosine or guanine opposite the 5'-thymine. (C) This fidelity of incorporation is altered for the polymerase domain of PrimPol lacking the zinc finger domain. Both Human and X. tropicalis PrimPol polymerase domains can incorporate an adenine or cytosine nucleotide opposite the 3'-thymine with approximately equal proficiency. If an adenine is incorporated in opposite the first base, the polymerase domain will then incorporate either a cytosine or thymine opposite the 5'-thymine of the (6-4)PP. If a cytosine is incorporated first, PrimPol$_{1\text{-}354}$ and XPrimPol$_{1\text{-}334}$ will incorporate adenine, cytosine or thymine opposite the 5'-thymine of the (6-4)PP.

this altered fidelity was orthologue-specific, we assayed the equivalent *Xenopus* polymerase domain ($XPrimPol_{1-334}$) and observed a similar fidelity to the human PrimPol polymerase domain (Figure 4.8C).

4.7.3. PrimPol Can Read Through an 8-oxoguanine Base

Next, we examined TLS bypass of an oxidative lesion, 8-oxo-G. In common with wild-type enzyme, $PrimPol_{1-354}$ can read through 8-oxo-G lesions with minimal stalling and exhibited the ability to insert A and C opposite 8-oxo-G lesions with equal efficiency (Figure 4.9A) (Bianchi et al., 2013; Engelke et al., 1990). A similar TLS activity and fidelity for this lesion was also evident with $XPrimPol_{1-334}$ (Figure 4.9A). This was not, however, the case for the Klenow fragment of the replicative Taq polymerase, which showed significant stalling at the 8-oxo-G lesion (Figure 4.9A). Upon testing the fidelity of PrimPol when incorporating opposite 8-oxo-G, we find that PrimPol has approximately equal efficiency in incorporation of dA and dC opposite this lesion, but Taq polymerase is only capable of inserting the Hoogsteen pairing base dA. This reorientation of the incoming base to form a Watson-Crick base pair with an 8-oxo-G lesion is evident in a number of other TLS polymerases, including pols η, ι, κ and REV1 (Haracska et al., 2000; Y. Zhang et al., 2002; 2001; 2000).

4.7.4. PrimPol is Unable to Read Through a Thymine Glycol Base

Thymine glycol lesions are, like 8-oxo-G, common products of oxidative damage to DNA. Neither wild-type, nor the polymerase domain of PrimPol harbour the ability to read through a thymine glycol lesion (Figure 4.10). To check whether this was orthologue-specific or whether other PrimPols might be able to read through a thymine glycol, we confirmed that this is not specific to human PrimPol using $XPrimPol_{1-334}$ (Figure 4.10). Like PrimPol, replicative polymerases, such as Taq polymerase, are also unable to read through a thymine glycol base (Figure 4.10). Pol η is capable of reading through a thymine glycol lesion, though its ability to extend from a thymine glycol lesion is dependent on the geometric isomer state of the thymine glycol lesion (Kusumoto et al., 2002).

4.7.5. PrimPol can Read Through Uracil Bases in DNA

Cytosine in DNA is susceptible to spontaneous deamination to uracil (dU) *in vivo*, which can base pair to adenine, resulting in C→T transition mutations (Duncan and

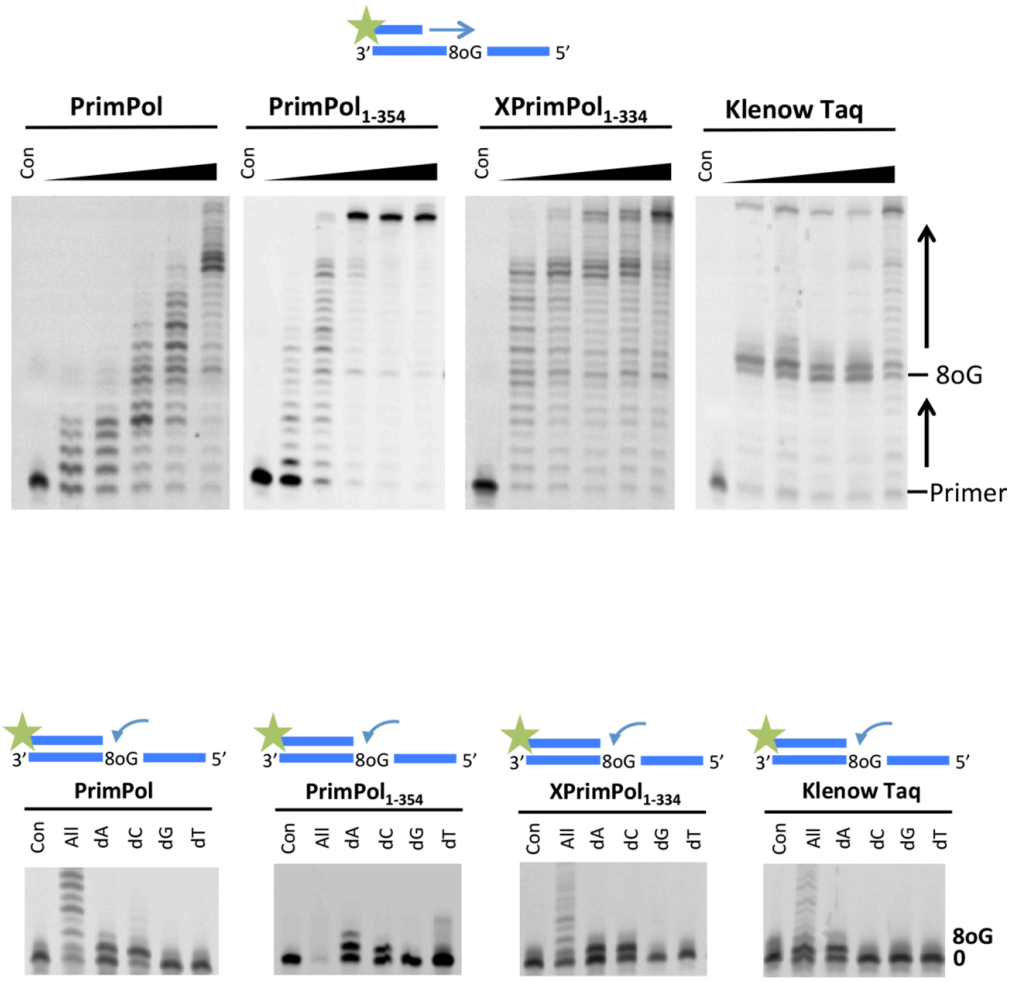


Figure 4.9. PrimPol can read-through an 8-oxoguanine lesion.
(A) PrimPol, $PrimPol_{1\text{-}354}$ and $XPrimPol_{1\text{-}334}$, as well as the replicative Klenow Taq polymerase, were incubated with dNTPs and a substrate with an 8-oxo-G (denoted 8oG above) lesion in its template (substrate 36 from Table 2.6) at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. PrimPol, $PrimPol_{1\text{-}354}$ and $XPrimPol_{1\text{-}334}$ are all capable of reading through this lesion with very little stalling, but Klenow Taq polymerase demonstrates an apparent stalling at the 8-oxo-G lesion. **(B)** These proteins were also assessed for their fidelity in incorporation opposite an 8-oxo-G lesion (substrate 37 from Table 2.6). Whilst Klenow Taq polymerase is only capable of inserting an adenine nucleotide opposite the 8-oxo-G lesion, PrimPol inserts both adenine and cytosine with approximately equal proficiency.

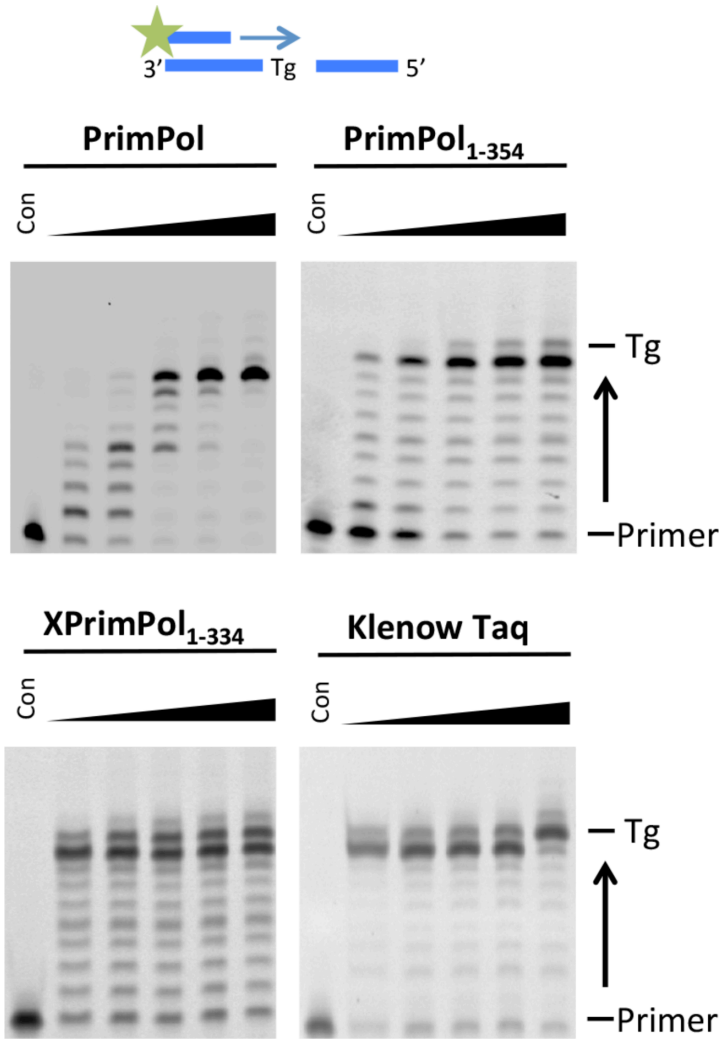


Figure 4.10. PrimPol cannot read-through a thymine glycol lesion.
PrimPol, PrimPol$_{1-354}$ and XPrimPol$_{1-334}$, as well as the replicative Klenow Taq polymerase, were incubated with dNTPs and a substrate with a thymine glycol lesion (denoted Tg above) in its template (substrate 38 from Table 2.6) at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. None of these proteins are capable of reading through a thymine glycol lesion.

Miller, 1980). Uracil is not recognised by, nor does it stall, eukaryotic replicative or TLS polymerases (Bocquier et al., 2001; Connolly et al., 2003; Lao-Sirieix and S. D. Bell, 2004; L. Liu et al., 2001). Similarly, the polymerase domain of PrimPol does not stall when confronted with a dU base (Figure 4.11). PrimPol incorporates an adenine opposite dU therefore, as one might expect, it essentially reads the dU base as a thymine. This is in common with $XPrimPol_{1-334}$ and the Klenow fragment of Taq polymerase (Figure 4.11).

4.7.6. PrimPol Cannot Read Through Apurinic/Apyrimidinic Sites

Apurinic/Apyrimidinic (AP) sites are endogenous DNA damage sites caused by spontaneous depurination, removal of uracil by uracil DNA glycosylase or as an intermediate in base excision repair (Section 1.4.5). Neither wild-type human, nor the polymerase domain of either human or *Xenopus,* PrimPol harbour the ability to read through an AP site (Figure 4.12). Whilst PrimPol is unable to bypass an AP site, there are a number of other polymerases that are capable of doing so. Incorporation of a nucleotide opposite an AP site is will likely be mutagenic as there is no way of knowing which nucleotide was there prior to removal. REV1 and pol ι can insert a single base opposite an AP site, which is subsequently extended by another TLS polymerase, pol ζ (Prakash et al., 2000; Y. Zhang et al., 2002). Pol κ can both insert a nucleotide opposite, and extend from, an AP site but the nucleotide inserted is dependent on the nucleotide in the primer immediately 5' of the lesion (Y. Zhang et al., 2000).

4.8. PrimPol can Incorporate dUTP Nucleotides

The presence of uracil bases in genomic DNA is not only the result of deamination of cytosine, but can also be caused by the insertion of free deoxyuridine triphosphate (dUTP) in place of dTTP. The essential enzyme Dut1 normally suppresses the incorporation of dUTP into genomic DNA (Gadsden et al., 1993), however the levels of dUTP in the dNTP pool are elevated during transcription and become available to insertion during genomic replication and/or repair synthesis (N. Kim and Jinks-Robertson, 2009). Whilst uracil itself is not promutagenic, as a polymerase will correctly insert an adenine opposite a uracil, the uracil base will be removed by uracil DNA glycosylase and this will result in an AP site. We aimed to determine whether PrimPol is able to utilise dUTP by measuring its incorporation opposite oligonucleotides containing different templating bases. PrimPol was incubated for 30 minutes in the presence of dUTP and four substrates containing the four different templating bases

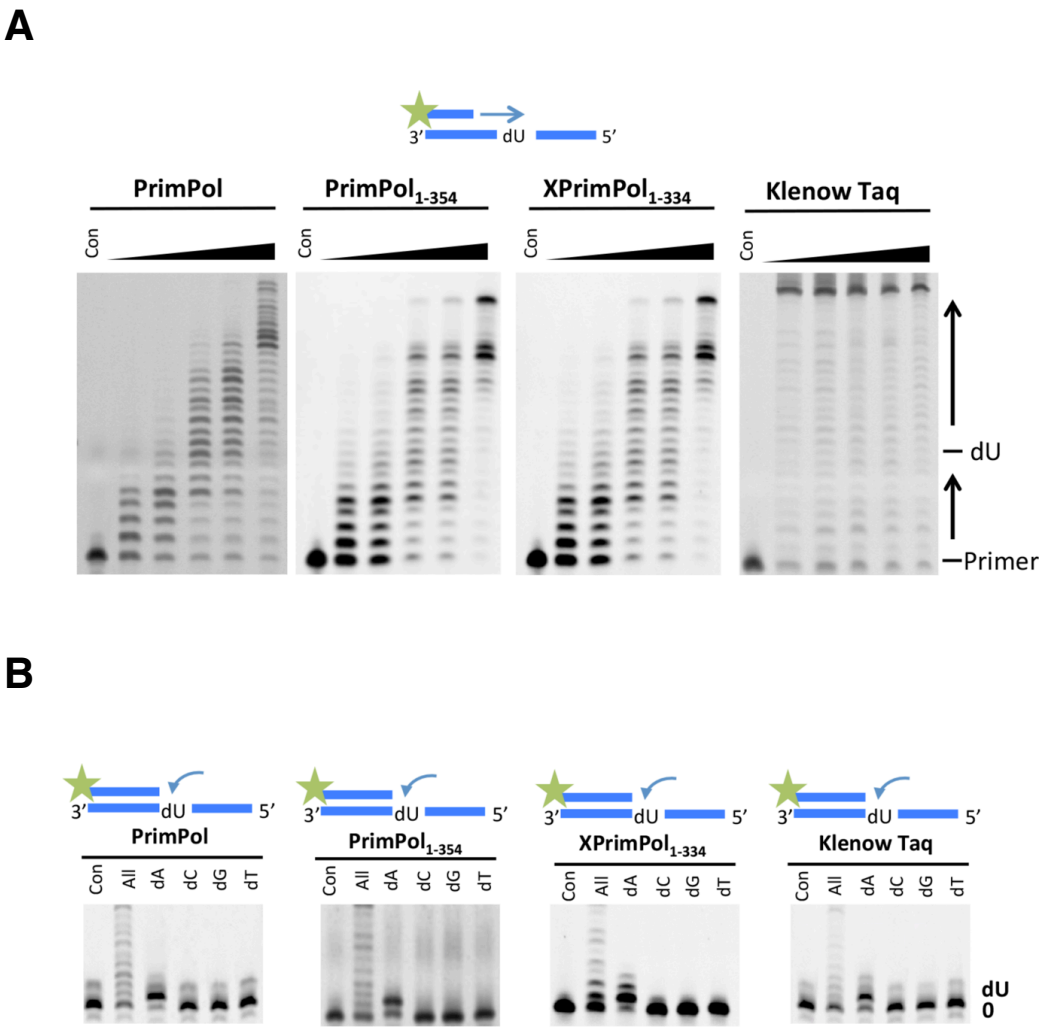


Figure 4.11. PrimPol interprets a uracil base in DNA as it does a thymine base. **(A)** PrimPol, PrimPol$_{1\text{-}354}$ and XPrimPol$_{1\text{-}334}$, as well as the replicative Klenow Taq polymerase, were incubated with dNTPs and a substrate with a deoxyuracil (denoted dU above) in its template (substrate 39 from Table 2.6) at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. All of these proteins will read through a uracil base in DNA without stalling. **(B)** Each of the proteins were assessed for their base incorporation activity opposite a uracil base. All proteins tested incorporate an adenine nucleotide opposite uracil, representing mutagenic activity if the source of the uracil base is cytosine deamination.

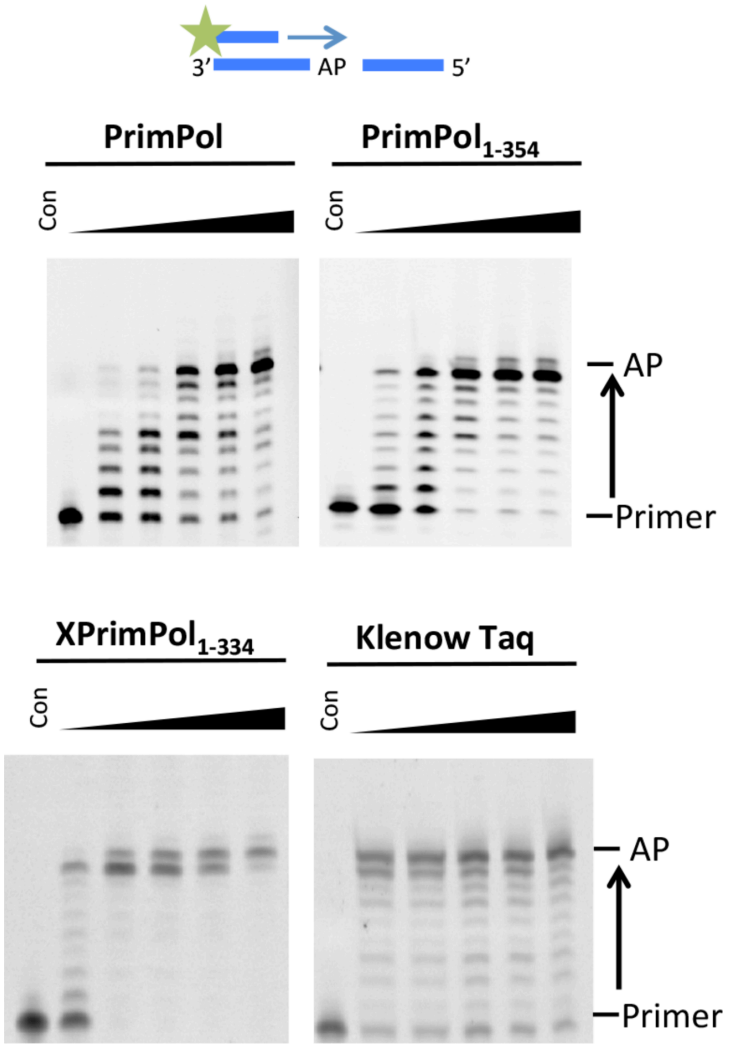


Figure 4.12. PrimPol cannot read-through an apurinic/apyrimidinic site.
PrimPol, $PrimPol_{1-354}$ and $XPrimPol_{1-334}$, as well as the replicative Klenow Taq polymerase, were incubated with dNTPs and a substrate with an AP site in its template (substrate 38 from Table 2.6) at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. None of these proteins are capable of reading through a site that doesn't have a templating base, indicating that an AP site is a significant replicative hindrance.

adenine, cytosine, guanine or thymine. We find that PrimPol is indeed able to incorporate uracil opposite an adenine template but is unable to do so opposite any other templating base (Figure 4.13), this is also the case with the polymerase domain $PrimPol_{1-354}$ (Figure 4.13).

4.9. PrimPol cannot Incorporate 8-oxo-dGTP Nucleotides

8-oxoguanosine triphosphate (8-oxo-dGTP) is a mutagenic DNA synthesis susbstrate formed by oxidation of dGTP in the dNTP pool as well as through the phosphorylation of 8-oxoguanosine diphosphate (8-oxo-dGDP) by nucleoside diphosphate kinase (Hayakawa et al., 1995; Maki and Sekiguchi, 1992). 8-oxo-dGTP is metabolised into 8-oxoguanosine monophosphate (8-oxo-dGMP) by 8-oxo-GTPase, a product which is unable to be rephosphorylated by nucleoside diphosphate kinase; this product is then itself further degraded by a specific nucleotidase to 8-oxoguanosine (8-oxo-G) (Hayakawa et al., 1995). We aimed to investigate whether PrimPol is able to incorporate 8-oxo-dGTP opposite different templating bases and found that PrimPol has an inability to incorporate this oxidised base opposite any of the four templating bases (Figure 4.14).

4.10. PrimPol's Zinc Finger Domain is Required for Bypass of Lesions *In Vivo*

To examine the ability of the different PrimPol variants to support bypass of bulky lesions during replication *in vivo*, a large pulse of UV irradiation (20 J/m^2) was introduced to cells between the two labelling steps to investigate fork stalling induced at UV lesions. This work was carried out in collaboration with my colleague Dr. Laura Bailey. PrimPol knockout DT40 cells have previously been shown a pronounced stalling phenotype, indicated by a large increase in the ratio of the CldU to IdU ratios following UV irradiation (Bianchi et al., 2013). Consistent with previous findings, we observe complementation by supplementing with the wild type enzyme (Figure 4.15). However, neither $PrimPol_{1-354}$ nor $PrimPol_{ZF-KO}$ were able to correct this fork stalling defect (Figure 4.15). Therefore, although the polymerase extension activity of PrimPol is sufficient for maintenance of unperturbed replication, these data suggest that its primase activity is requisite for timely replication fork progression following the introduction of significant DNA damage.

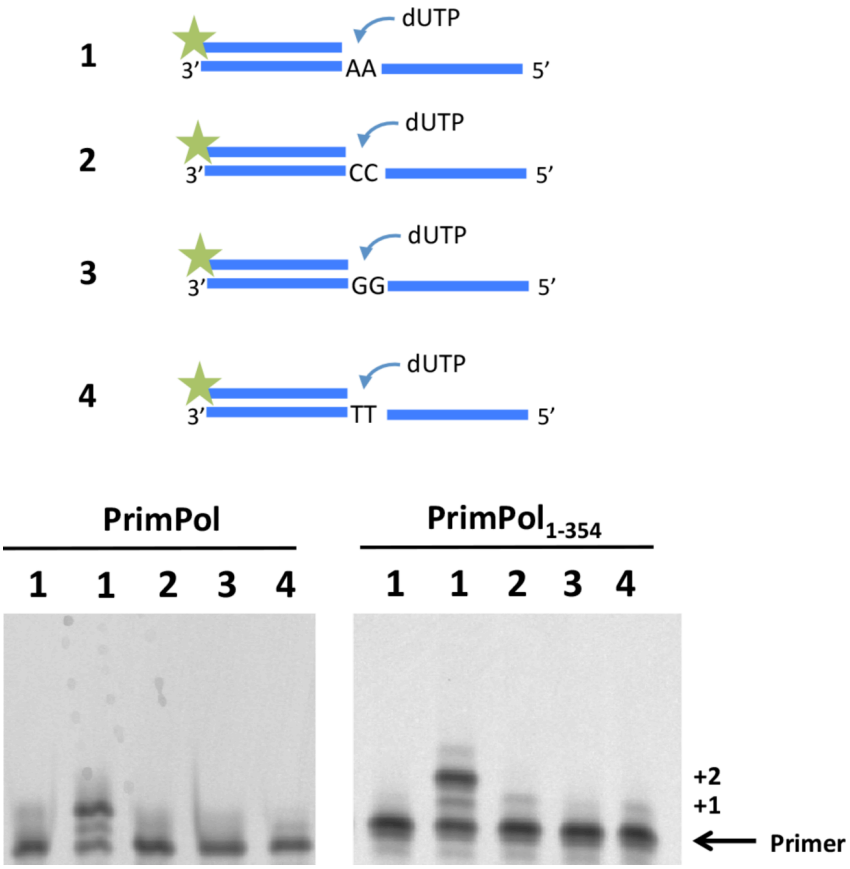


Figure 4.13. PrimPol is capable of incorporating dUTP into DNA.
As cytosine can be deaminated in the cellular nucleotide pool, PrimPol and PrimPol$_{1\text{-}354}$ were tested for their ability to incorporate deoxyuracil triphosphate (dUTP) into DNA opposite four different templating bases (substrates 9-12 from Table 2.6), alongside a control in which no protein was added. dUTP is incorporated proficiently opposite adenine but not any other base. This is to be expected as dUTP readily base pairs with adenine and PrimPol will insert an adenine opposite a uracil base in template DNA (Figure 4.11B).

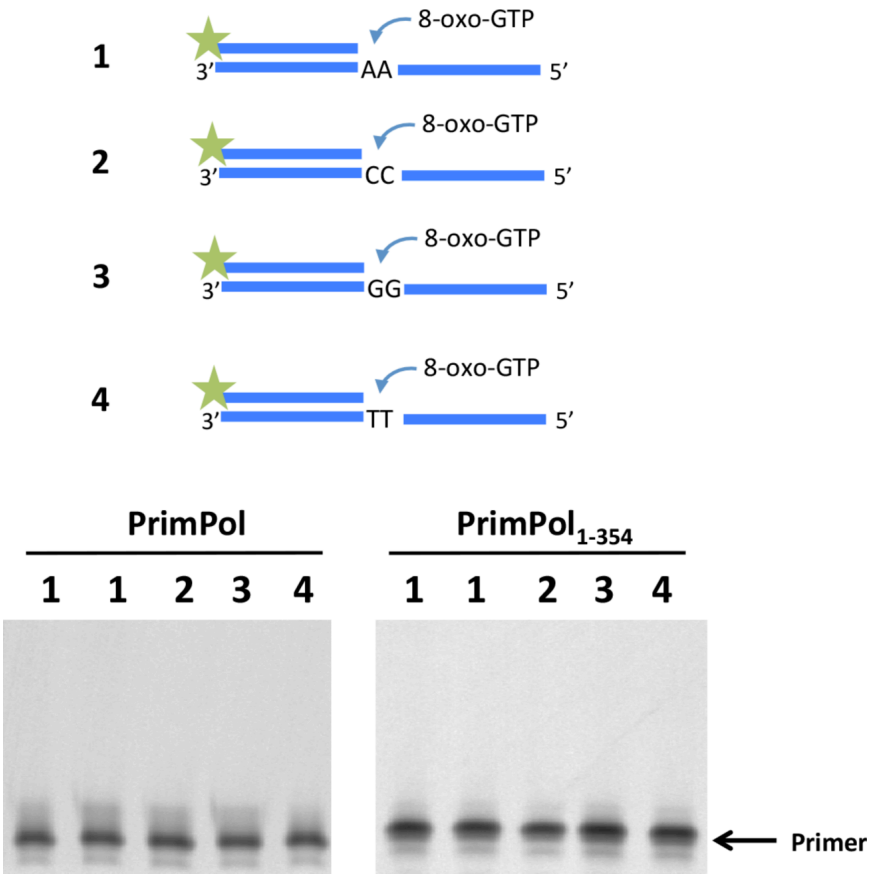


Figure 4.14. PrimPol is not capable of incorporating 8-oxo-dGTP into DNA. As guanine is liable to oxidation in the cellular nucleotide pool, PrimPol and PrimPol$_{1\text{-}354}$ were also tested for their ability to incorporate 8-oxoguanosine tripohosphate (8-oxo-GTP) into DNA opposite four different templating bases (substrates 9-12 from Table 2.6), alongside a control in which no protein was added. Neither PrimPol nor PrimPol$_{1\text{-}354}$ are capable of inserting 8-oxo-GTP into DNA. This is in contrast to PrimPol's ability to incorporate adenine and cytosine bases opposite a templating 8-oxo-G in DNA (Figure 4.#B), which suggests that the ability to accommodate a lesion in the DNA substrate and the ability to accommodate a lesion in the incoming nucleotide substrate are not homogeneous.

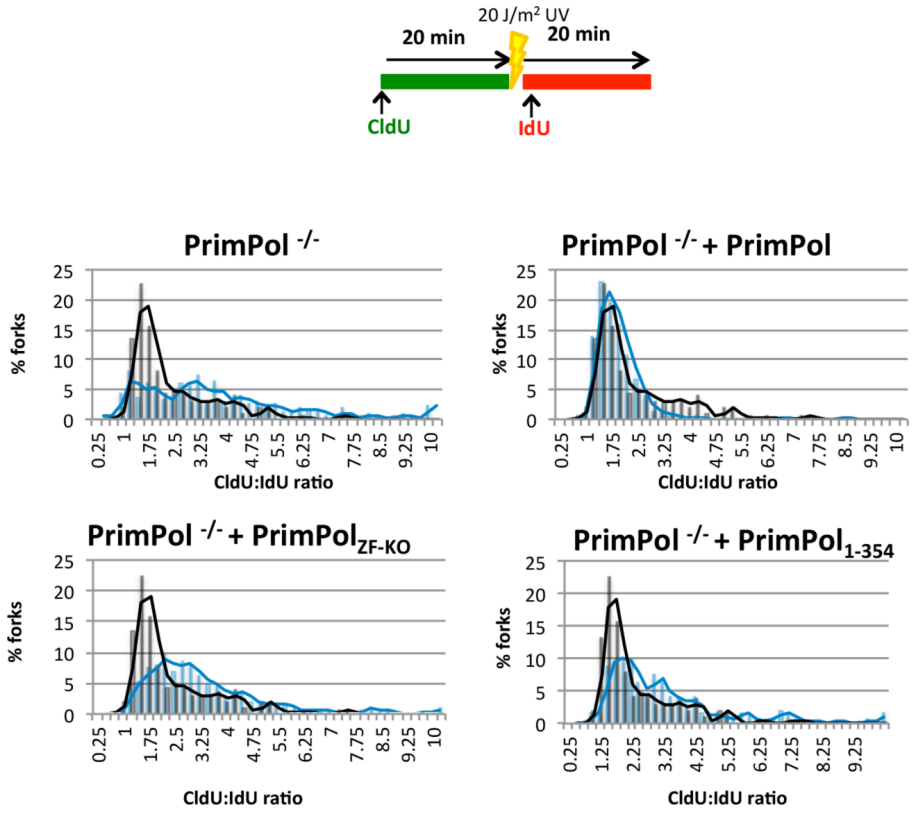


Figure 4.15. The zinc finger of PrimPol is required for maintaining fork rate following UV irradiation.
DNA fork rate analysis was carried out with after a 20 J m^{-2} pulse of UV, between CldU and IdU labels, to look for fork-stalling. Data is shown as the ratio of the two labels, Cldu, pre-UV : IdU, post UV label (n = 3). The black lines indicate the Cldu:IdU ratio of the two labels in the wild-type cells and the blue lines indicate the ratios in PrimPol$^{-/-}$ cells complemented with the indicated PrimPol variants. Wild-type PrimPol is able to successfully complement the PrimPol$^{-/-}$ cells but the PrimPol$_{ZF\text{-}KO}$ and PrimPol$_{1\text{-}354}$ variants, lacking an intact zinc finger, are unable to complement fork rate following UV irradiation. The requirement for the zinc finger indicates that the primase activity of PrimPol is required for maintaining fork rate following UV irradiation.

4.11. Summary and Discussion

In this chapter, we have probed the polymerase activities of PrimPol and drawn comparisons with other polymerases. We have also compared the wild-type enzyme to the catalytically active polymerase domain of human PrimPol; although the DNA polymerase activity of PrimPol appeared initially to be independent of the zinc finger domain, PrimPol polymerase domain lacking the C-terminal zinc finger domain ($PrimPol_{1-354}$) showed enhanced relative polymerase and processivity activities, compared to the wild-type enzyme. Further studies revealed that the zinc finger modulates the fidelity and extension activities, including TLS activities, of PrimPol and therefore it should also be considered to be an important regulator of PrimPol's polymerase functions as well as requisite for primase activity.

PrimPol is a lower fidelity polymerase than replicative polymerases. As such, it is required that this protein is regulated in its activity; else it would replicate long tracts of DNA in an error-prone manner. PrimPol is a very distributive enzyme, catalysing up to 4 subsequent incorporations during a single DNA substrate binding event/DNA polymerisation cycle *in vitro*. This distributive character of PrimPol is reminiscent of the DNA polymerase processivity reported for the Y-family TLS polymerases, such as Pol κ and η, which predominantly incorporate up to 6 or 7 nucleotides (Ohashi et al., 2000; Washington et al., 1999). It cannot be ruled out, however, that PrimPol processivity may be enhanced through the binding of accessory proteins.

As we see the same type of substitution mutation from $PrimPol_{1-354}$ here as is evident in the gapped-plasmid assay that uses wild-type PrimPol (Guilliam et al., 2015), this indicates that the lack of zinc finger does not necessarily alter the fidelity of the polymerase domain, rather exacerbate the inherent poor fidelity of PrimPol's polymerase. Additional kinetic analysis is required to determine whether the observed fidelity is primarily a result of differences in the incorporation rates for correct versus incorrect bases rather than differences in nucleotide binding affinity. Also, given that PrimPol is highly selective for insertions and deletions (Guilliam et al., 2015), the incorporation of a third nucleotide opposite an incorrect templating base by both wild-type PrimPol and $PrimPol_{1-354}$ could actually be the result of PrimPol scrunching the primer and using the same templating base more than once in order to adapt and continue extension.

In the presence of manganese, PrimPol appeared to be much more active than with magnesium. However, upon closer inspection, we discovered that the enzyme can produce polymeric primer extension products, even when only a single form of dNTP is present. Under these manganese-dependent conditions, PrimPol extends from a primer to form homopolymeric strands that are non-complementary and therefore unlikely to be anneal to the template strand. This suggests that caution should be taken when interpreting the results from studies using manganese as the resulting products may represent aberrant synthesis products rather that the expected template-dependent ones. Manganese has the effect of decreasing fidelity amongst polymerases (Hays and Berdis, 2002; Tabor and Richardson, 1989). Whilst magnesium enforces tetrahedral geometry in the arrangement of its ligands, manganese will accommodate square, planar, tetrahedral and octahedral co-ordinations thus increasing the ability to accelerate the rate of reaction in substrates that are perhaps misaligned (Pelletier et al., 1996). The physiological implications of this template-independent primer extension activity of PrimPol in the presence of manganese remains unclear, as the cellular concentrations of manganese are low, suggesting this activity may be irrelevant *in vivo* or it may be limited to certain cellular compartments (e.g. mitochondria).

As PrimPol can incorporate bases opposite, and extend from, highly distorting lesions it is likely that the active site involved in production of phosphodiester bond is large and flexible to accommodate a range of lesions. The (6-4)PPs, in particular, are highly mutagenic due to the high level of distortion they introduce into the DNA backbone. The helical bending at these lesions is estimated to be ~44° and the distance between phosphorus atoms in the paired bases in the (6-4)PP is ~6Å, compared to to the 7Å average corresponding distance at other base pairing steps (J.-K. Kim and Choi, 1995). Therefore, it is probably unsurprising that PrimPol exhibits low fidelity when inserting bases opposite a (6-4)PP, as a polymerase that can accommodate this lesion would require a flexible active site. Unexpectedly, the polymerase domain of PrimPol, like Pol η, is capable of bypassing CPDs with high fidelity. This contrasts with full length PrimPol, which is not capable of bypassing CPDs unless two adenines are inserted opposite the CPD (Bianchi et al., 2013; Rudd et al., 2013). This result establishes that the active site of PrimPol has the capacity to bypass these lesions but this is constrained in the full-length enzyme. A conformation change may be required,

potentially induced by binding partners that alters the activity of PrimPol to allow it to bind to and bypass CPDs *in vivo.*

PrimPol is also confronted with a number of lesions as a result of oxidative stress in cells. These include apurinic/apyrimidinic (AP Sites) and 8-oxo-G lesions. Pol γ, the mitochondrial replicative polymerase, has the ability to bypass an 8-oxo-G site but at a reduced rate, and cannot bypass AP site (Graziewicz et al., 2007; Pinz et al., 1995). Conversely, PrimPol shows almost no hindrance to base incorporation opposite 8-oxo-G lesions and, although it cannot incorporate bases opposite an AP site, it has the capacity to re-prime post-lesion. 8-oxo-G is expected to pair with C through Watson-Crick base pairing in its regular *anti* conformation, but in its *syn* conformation it forms Hoogsteen base pairs with adenine (van Loon et al., 2010). This ability to insert A, as well as C, opposite 8-oxo-G is apparent in a number of polymerases but the addition of auxiliary proteins (e.g. RPA and PCNA) can dramatically increase the preference for adenine over cytosine, as is the case with Pols λ and η (Maga et al., 2007; van Loon et al., 2010). It remains to be seen whether this is also the case with PrimPol. This ability to accommodate bulky lesions in the active site comes at the expense of fidelity and, therefore, PrimPol is a low fidelity polymerase that is likely to make limited and, perhaps, some nonspecific contacts with the replicating base pair. Deoxyuracil is not recognised as a lesion by replicative polymerases and results in C:G → T:A transition mutations. Deoxyuracil is recognised by archaeal family B polymerases as a lesion and stalls replication to prevent mutation (Fogg et al., 2002) but PrimPol does not have this recognition capacity and incorporates an adenine opposite a uracil base.

It is probable, given the activities of PrimPol and its phyletic profile, that PrimPol was acquired early in eukaryotic evolution to play a significant role in lesion bypass, particularly of oxidative and UV lesions, during DNA replication. PrimPol-mediated damage tolerance appears to represent an important mechanism that facilitates replication fork restart following stalling of the replisome at sites of DNA damage.

Intriguingly, because PrimPol is able to incorporate an undamaged dNTP opposite an 8-oxo-G template but is unable to incorporate 8-oxo-GTP opposite an undamaged template base. This suggests that although the active site of PrimPol is accommodating on the template side, it is less accommodating to incoming damaged bases. PrimPol is able to incorporate a dUTP nucleotide because it is not particularly

bulky; it has fewer side chains than dCTP and readily forms Watson-Crick base pairs with adenine.

Canonical DNA polymerases share a common right-handed architecture with the polymerisation domain containing fingers, palm and thumb subdomains. All of the three structural elements are clustered together providing accurate, highly structurally restrained positioning of the polymerase at the primer-template junction in order to promote the highest fidelity and processivity of DNA synthesis. Although this tight and very specific binding cleft is perfectly suited for replication of entire genomes, in some instances such structural arrangement can be disadvantageous. In particular, it limits the capacity of the replicative enzymes to tolerate abnormalities in the topology of DNA templates, such as secondary structures or DNA damage-induced distortions. The thumb subdomain of replicative polymerases interacts with the sugar-phosphate backbone of the DNA to coordinate the DNA. This subdomain is optimised for the highly accurate replication of DNA and is sensitive for structural anomalies in the in the DNA template. The predicted structure of PrimPol, in addition to the known structures of primases, suggests that this polymerase lacks structural elements corresponding to the 'thumb' domain that is present in replicative polymerases that is implicated in processivity (Brautigam and Steitz, 1998). The open planar predicted structure of this polymerase indicates a transient binding of the polymerase domain and this is consistent with other low fidelity polymerases, such as those of the Y-family of polymerases (Lehmann et al., 2007).

The processivity data presented here suggest that the zinc finger domain of PrimPol regulates the fidelity of these polymerases, in addition to its essential role in primase activity. Whilst there is no reduction in the activity of the polymerase domain lacking the zinc finger, it actually appears more active, we propose that the presence of the zinc finger reduces the processivity of the enzyme to allow a slower, higher fidelity incorporation of complementary nucleotides. We postulate that whilst the AEP polymerase domain provides a platform to recognise the primer-template junction of the DNA substrate, the zinc finger binds the single-stranded region of the template strand to modulate binding and fidelity by the AEP domain. Y-family polymerases have a highly truncated thumb domain and, notably, contain an additional functional module referred to as the polymerase associated domain (PAD). Interestingly, studies on TLS polymerases Dpo4 and Dbh point to a role for PAD in regulation of processivity and

fidelity of these enzymes (Boudsocq et al., 2004). This structural element not only binds ssDNA and provides an extended flexible platform coordinating the template strand during DNA synthesis on damaged DNA templates, but also modulates the activity of these TLS DNA polymerases. As part of an adaptation to perform TLS, we suggest that the zinc finger is functionally analogous to the PAD subdomain, but is spatially separated to allow greater freedom of binding of the enzyme to distorted DNA templates, such as those containing UV-induced lesions.

Regarding the bypass of lesions during genome replication *in vivo*, this and previous work (Bianchi et al., 2013) suggests that both of PrimPol's priming and extension activities are requisite for lesion bypass and replication restart. When replicative DNA polymerases encounter a blocking lesion, they stall and PrimPol is recruited, possibly by RPA (see Chapter 6), to perform TLS or re-prime replication downstream of the lesion (Figure 4.16). After production of a primer or polymerisation through the lesion, the replicative polymerase can then resume replication.

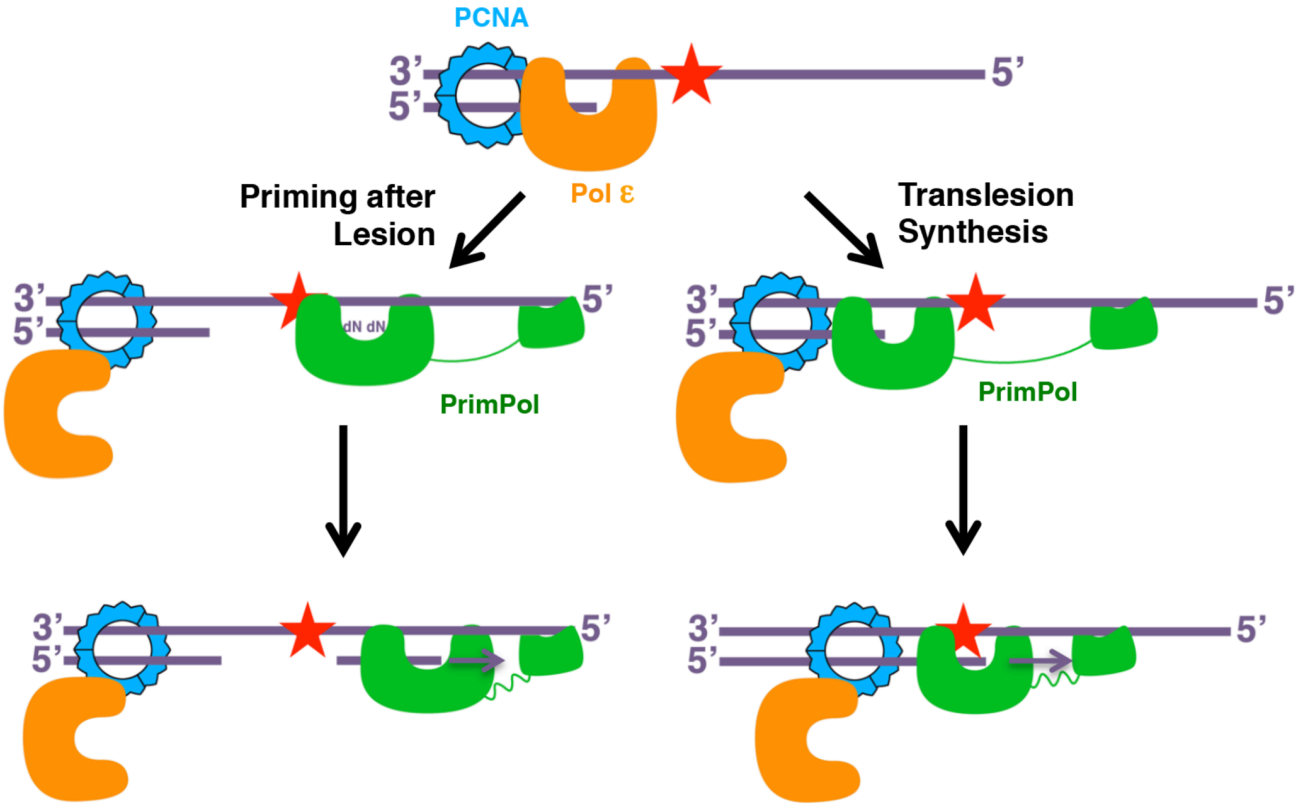


Figure 4.16. Model of the lesion bypass mechanisms of PrimPol.
PrimPol utilises both its primase and polymerase activities to bypass DNA damage that blocks replicative polymerases. When replicative polymerases encounter a blocking lesion, they are displaced or regress from the site of damage. PrimPol can then access the template upstream of the lesion, where it can reprime following the lesion (left) or if it can access the primer-template junction it can carry out TLS (right). Once PrimPol-mediated bypass of the lesion has occurred, the replicative polymerase can then resume replication at the fork.

CHAPTER 5

A Human PrimPol Mutation Associated with High Myopia has a DNA Replication Defect

5.1. Introduction

Although recent studies have established the importance of PrimPol in damage tolerance during DNA replication in cellular and animal model systems, its association with human diseases has only been conjectured. However, exome sequencing of the genomes of candidate patients has recently implicated the PrimPol gene as a possible susceptibility gene associated with high myopia (Zhao et al., 2013). This study identified a missense mutation in human PrimPol caused by a thymine to guanine transversion, resulting in the substitution of tyrosine at position 89 by an aspartic acid residue (Y89D). This mutation was not only found in a single family but also in four additional sporadic patients with high myopia and these patients had high visual refractive errors (-6.5 to -32.15 diopters, D) (Zhao et al., 2013). Myopia, or short-sightedness, is a common problem around the world in those with visual defects and this problem is rapidly increasing. Myopia accounts for 41% of decreased vision in Western populations and this up from 25% in the early 1970s (Vitale et al., 2009). In some Southeast Asian populations up to 80-90% of all children leaving school are now myopic (Morgan et al., 2012). 10-20% of these children suffer from high myopia, an extreme form of the disease, often characterised by a refractive error of less than -6.00 D and is often associated with other serious ocular disorders (Morgan et al., 2012; Saw et al., 2005). The aetiology of high myopia remains extremely complex and probably involves the synergy between many genes.

The active site of PrimPol is predicted to consist of three magnesium-coordinating, negatively-charged residues: an aspartate at residue 114, a glutamate at residue 116 and an aspartate at residue 280. Although the tyrosine at position 89 is not predicted to reside directly within the active site, it is in close proximity and therefore may play yet undetermined roles in the activities associated with these enzymes. The mutated tyrosine (Y89) is highly conserved residue across a wide array of chordate species (Figure 5.1) and *PolyPhen-2* prediction programme indicates that mutation of this residue would be functionally damaging (Zhao et al., 2013). In addition to PrimPol, this tyrosine is also present in the related predicted primase PF14_0050 in *plasmodium falciparum*, so its evolutionary conservation is indicative of a significant functional residue (L. M. Iyer et al., 2005).

In this study, we aimed to characterise the biochemical properties of the Y89D mutant of PrimPol (PrimPol^{Y89D}) to determine if this amino acid change alters PrimPol's

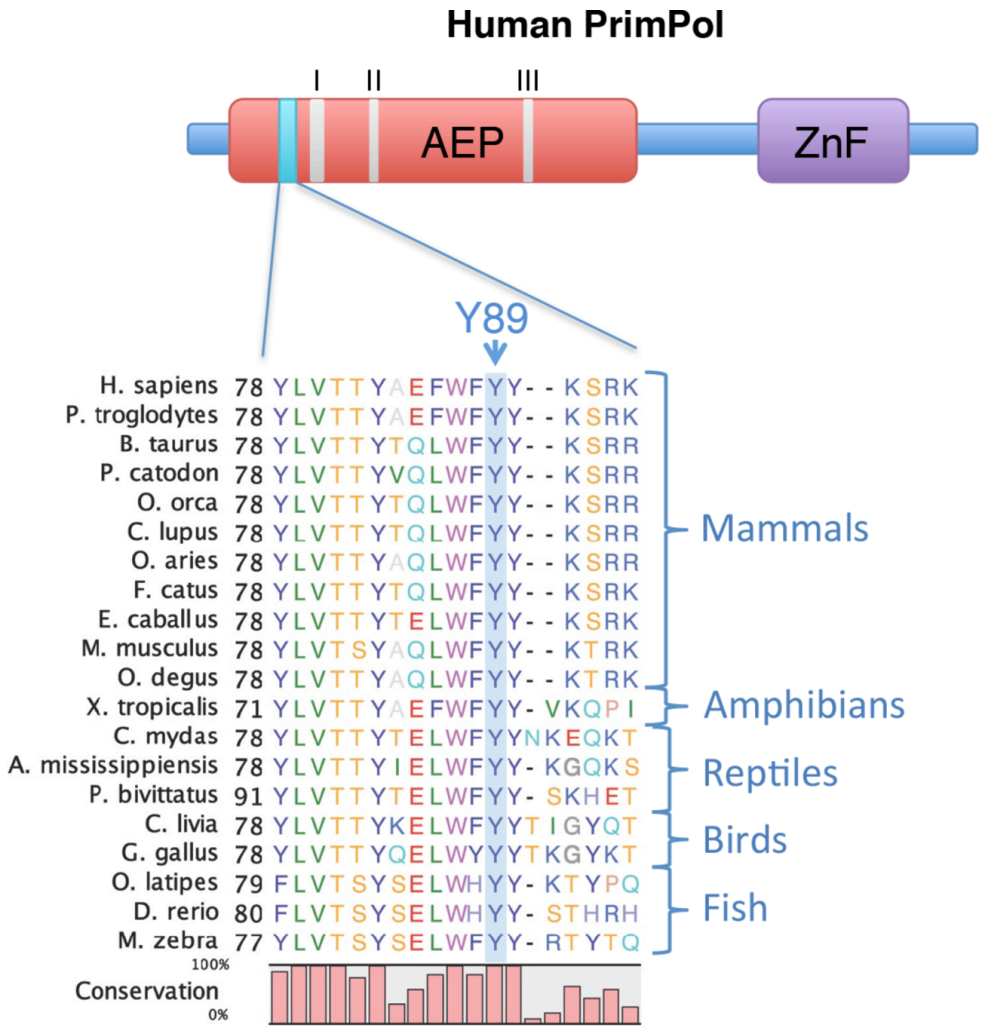


Figure 5.1. Tyr89 is a highly conserved residue amongst higher eukaryotic species.
The tyrosine at position 89 in human PrimPol is situated in the AEP polymerase domain of PrimPol and this residue is highly conserved among higher eukaryotes across mammalian, amphibian, reptilian, avian and piscine species.

activities thus establishing a molecular basis for its proposed involvement in high myopia. Here, we demonstrate that the Y89D mutant loses its ability to prime using rNTPs, but can still prime using dNTPs. PrimPol^{Y89D} has a greatly reduced template-dependent processivity activity, when compared with the wild-type enzyme. In addition, we report that this mutant has significantly reduced binding for both DNA and dNTPs but its fidelity remains largely unchanged. Furthermore, we show that it is the hydrophobic ring of the tyrosine at position 89 that is important for maintaining these activities. We also demonstrate that changes in enzyme activity are the result of global changes in the structure of the polymerase domain of PrimPol. Finally, we report that this mutation significant slows replication fork rates *in vivo* and reduces cell viability after DNA damage. Together, these data establish that a point mutation identified in PrimPol from patients with high myopia results in a major disruption of the catalytic and replication activities associated with human PrimPol thus establishing a potential link between replication stress and human disease, particularly high myopia.

5.2. Construction and Purification of PrimPol Point Mutants

To probe the altered activities of the Y89D variant of PrimPol this point mutant and number of others were constructed and purified. The PrimPol$_{Y89D}$ construct was produced by site-directed mutagenesis (plasmid H13, Table 2.3) and purified by Dr. Stanislaw Jozwiakowski. Two other variants of this point mutant, Y89S and Y89F, were both constructed (plasmids H15 and H16, Table 2.3) were also constructed by site directed mutagenesis and purified as described in Tables 2.4 and 2.5 (Figure 5.2A and B). The tyrosine side chain at position 89 consists of a hydrophobic aromatic ring and a hydrophilic hydroxyl group; to determine whether altered activities can be attributed to one or the other of these moieties, a serine with a hydrophilic hydroxyl group and a phenylalanine with a hydrophobic aromatic ring side chain were both constructed. Additionally, to compare any changes in the polymerase domain caused by this Y89D mutation, this mutant was also introduced into the polymerase domain of PrimPol (PrimPol$^{Y89D}{}_{1\text{-}354}$; plasmid H14, Table 2.3) and purified as described in Tables 2.4 and 2.5 (Figure 5.2C).

5.3. PrimPol^{Y89D} retains DNA primase Activity but not RNA Primase Activity

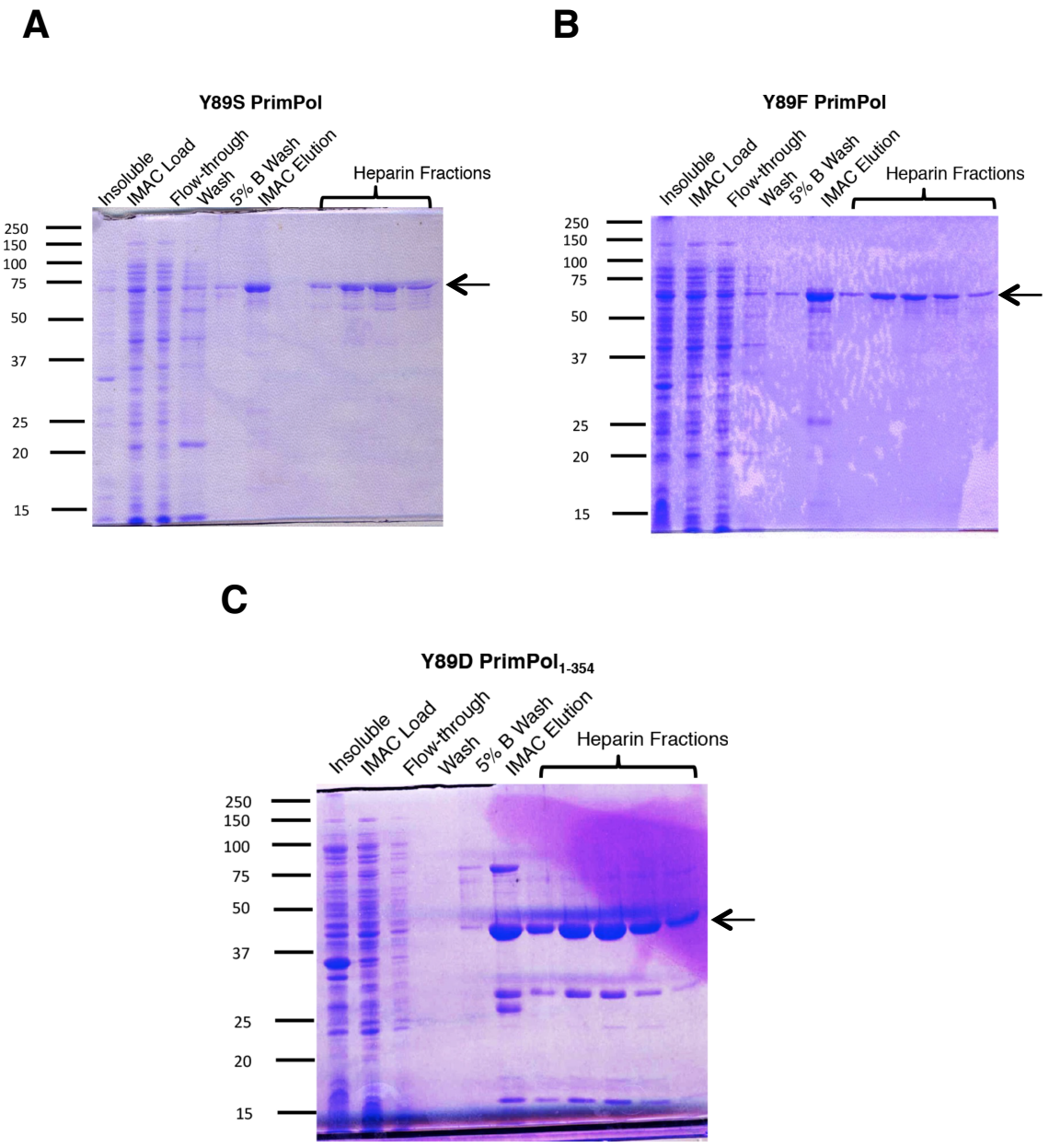


Figure 5.2. Purification of PrimPol^{Y89S}, PrimPol^{Y89F} and PrimPol$^{Y89D}_{1-354}$ variants. Samples from the insoluble and soluble (load) fractions from the growth and expression of proteins, as well as samples of the fractions from IMAC affinity purification and protein-containing heparin column elution fractions are shown for each of **(A)** PrimPol^{Y89S} (68.8 kDa) **(B)** PrimPol^{Y89F} (68.9 kDa) and **(C)** PrimPol$^{Y89D}_{1-354}$ (43.1 kDa).

To determine if PrimPol displays an underlying biochemical defect that contributes to the high myopia clinical phenotype, we first examined the primase activity of the PrimPol^{Y89D} variant. These experiments were carried out in collaboration with my colleague Dr. Stanislaw Jozwiakowski as described in Section 3.10. *De novo* primer synthesis remained largely unchanged between the mutant and the wild-type in the production of DNA primers from dNTPs (Figure 5.3). However, unlike the wild-type enzyme, PrimPol^{Y89D} was unable to synthesise RNA primers in the presence of even high concentrations of rNTPs. Whilst it has been reported that PrimPol is more proficient at producing DNA over RNA primers *in vivo*, we remained intrigued as to why this mutation in PrimPol conferred a defect in RNA primer synthesis. Upon testing PrimPol^{Y89S} and PrimPol^{Y89F} for primase activity (Figure 5.3), we found that the PrimPol^{Y89S} has an activity comparable to Y89D but the Y89F variant of PrimPol returned primase activity back to wild-type, restoring the ability of the enzyme to produce RNA primers. This implies that the hydrophobic aromatic ring moiety at position 89 is required for optimal PrimPol activity.

5.4. The Y89D Variant of PrimPol has Limited Processivity

PrimPol possesses two intrinsic catalytic activities (primase and polymerase), therefore to determine any catalytic deficiencies associated with the Y89D point mutation, we must also consider its polymerase activity in addition to its primase activity. Primer extension assays on PrimPol showed that the polymerase activity of PrimPol^{Y89D} was significantly reduced in comparison to the wild-type enzyme (Figure 5.4). We next tested if this activity was restored by either Y89S or Y89F variants of PrimPol (Figure 5.4) and observed that Y89S had slightly higher activity than the Y89D variant but was not restored to wild-type levels. In contrast, Y89F was fully proficient at extending primers opposite a templating strand. Notably, although PrimPol^{Y89F} was able to fully extend primers, and showed full extension at time points as short as 3 minutes (Figure 5.4, Bottom-right, Lane 4), more primers were left unextended compared with wild-type PrimPol, suggesting that this modified enzyme has considerable activity once it has bound DNA but coordinates it less efficiently.

PrimPol was previously shown to be a distributive DNA polymerase, inserting only up to 4 nucleotides during a single binding event (Figure 5.5). We next determined whether the processivity of the Y89D mutant was altered, causing the observed reduction in the polymerase activity. PrimPol^{Y89D} variant of could only insert a single

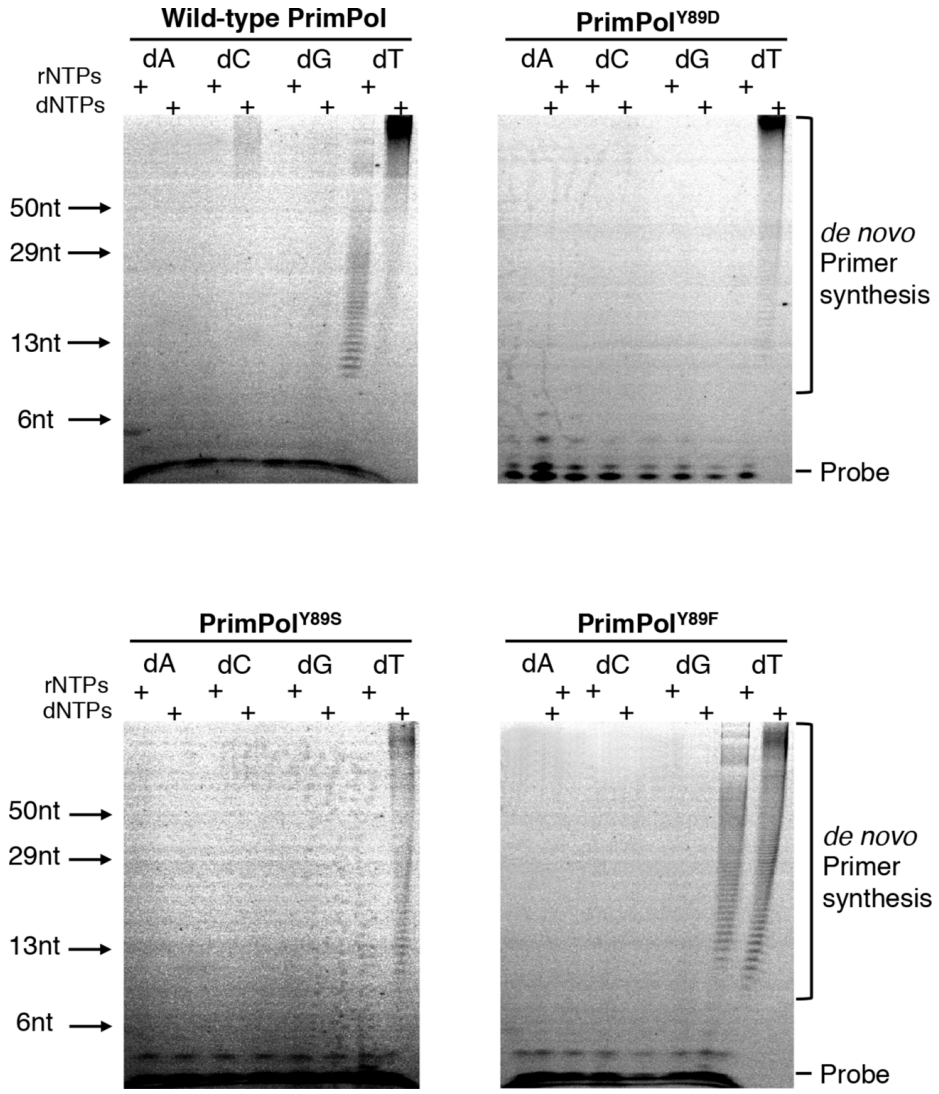


Figure 5.3. Primase activities of human PrimPol and PrimPol^{Y89D}.
Human PrimPol has primase activity and can produce *de novo* primers using rNTPs and dNTPs in a poly(dT) homopolymer template-dependent fashion. Substrates shown here are substrates 4-7 in Table 2.6. PrimPol^{Y89D} partially loses its primase activity. This variant of PrimPol cannot synthesize RNA primers but is an active primase utilising dNTPs. PrimPol^{Y89S} shows significant loss of the enzymatic activity with residual primase activity observed in presence of dNTPs. PrimPol^{Y89F} is an active primase variant capable of synthesizing both RNA and DNA primers with efficiency comparable to wild-type PrimPol. Experiments shown here were carried out by Dr. Stanislaw Jozwiakowski.

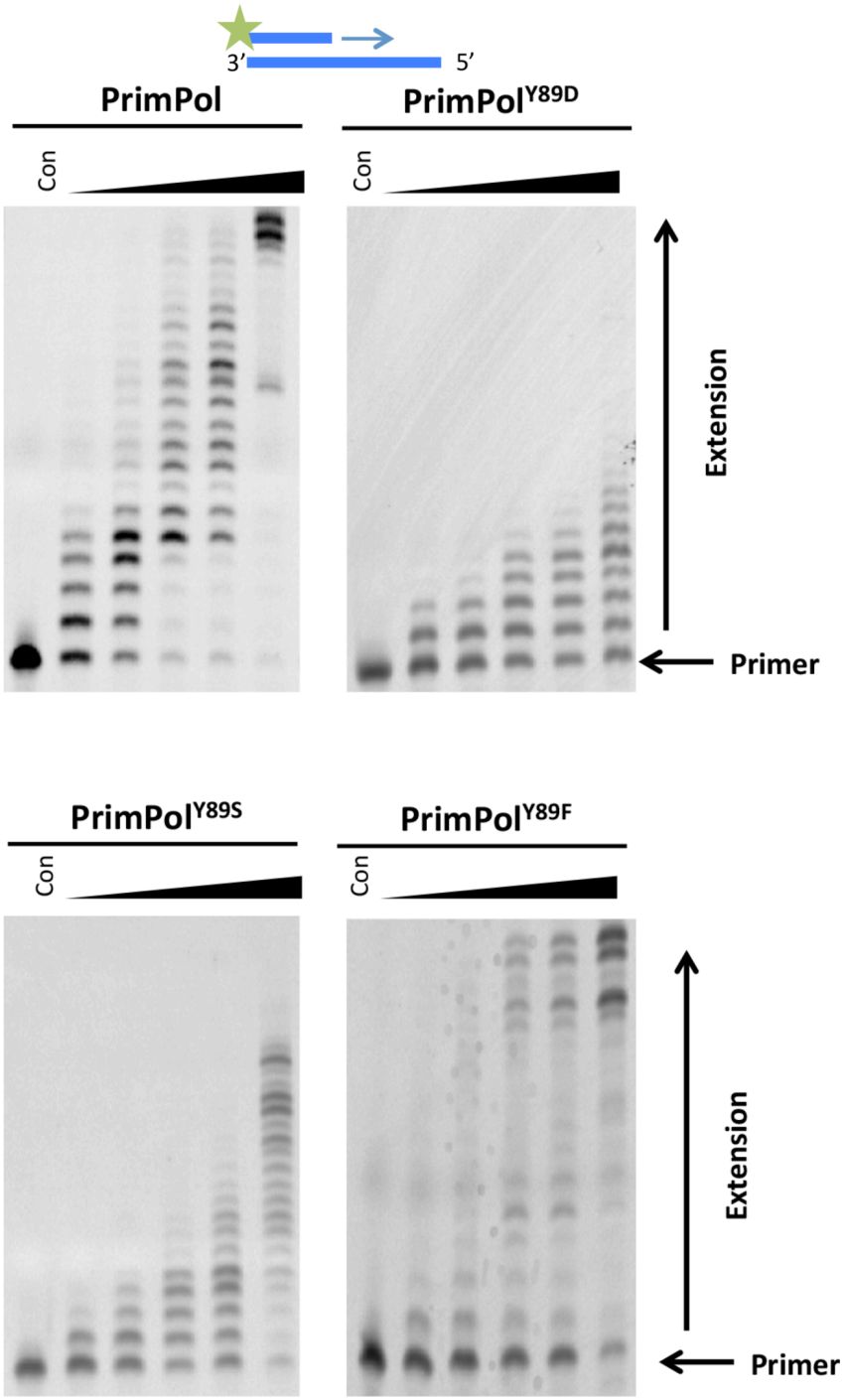


Figure 5.4. Primer extension activities of human PrimPol and PrimPol^{Y89D} .
Human PrimPol has primer extension activity and produces full-length templates following incubation with dNTPs and substrate 1 from Table 2.6 at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. PrimPol^{Y89D} shows a marked reduction in polymerase activity, incorporating up to eight nucleotides following incubation for 60 mins. PrimPol^{Y89S} also shows a loss in polymerase activity relative to wild-type PrimPol, though it is a more active polymerase than PrimPol^{Y89D}. PrimPol^{Y89F} restores polymerase activity. The Y89F variant can fully extend primers but there is an increased level of unextended primer relative to wild-type PrimPol.

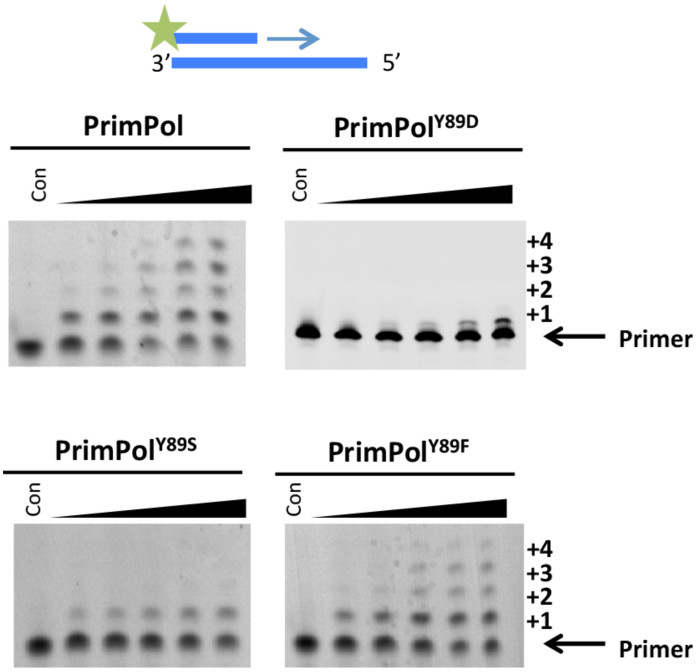


Figure 5.5. Processivity of human PrimPol and PrimPol^{Y89D}.
Primer extension assays were performed in the presence of an excess of herring sperm DNA trap to ensure the polymerase only binds once to the template DNA. The first lanes represent a no-enzyme control and lanes 2–6 represent time points of 15, 30, 60, 120 and 360 s. As demonstrated previously in Chapter 3, Human PrimPol is a poorly processive enzyme, inserting up to four nucleotides opposite a templating DNA strand in a single binding event. PrimPol^{Y89D} has marked reduction in processivity, and is only able to insert a single nucleotide opposite a templating DNA strand per single binding event. PrimPol^{Y89S} can also insert only one single nucleotide opposite a templating DNA strand. PrimPol^{Y89F} restores processivity and has the ability to insert up to four nucleotides, similarly to wild-type PrimPol.

nucleotide opposite a templating strand of DNA, making it an even more distributive enzyme than wild-type PrimPol (Figure 5.5). To determine whether the aromatic ring of tyrosine was again the main determinant for enzyme activity, we tested the processivity activities of Y89S and Y89F variants. Consistent with Y89D, PrimPol^{Y89S} could only insert a single nucleotide. However, in common with wild-type enzyme, PrimPol^{Y89F} could insert up to 4 nucleotides. This establishes that the presence of an aromatic side-chain at this position is requisite for maintaining more processive extension by PrimPol.

5.5. PrimPol^{Y89D} has Reduced DNA Binding Activity

A reduction in processivity of polymerases can be attributed to alterations in either dNTP or substrate DNA binding. To test the DNA binding efficiencies of wild-type and PrimPol^{Y89D}, we introduced the mutation into the polymerase domain (PrimPol$_{1\text{-}354}$) of PrimPol. This ensured that the contributory DNA binding effects of the zinc finger were removed to enable accurate assessment of the alteration in binding caused solely by this point mutation in the polymerase domain. As the polymerase domain is able to bind both single-stranded and double-stranded DNA (Section 3.7), we chose to use a labelled overhanging DNA substrate in our EMSAs similar to the substrates PrimPol would encounter *in vivo*. The binding efficiency of PrimPol$^{Y89D}{}_{1\text{-}354}$ to DNA was almost four times lower than PrimPol$_{1\text{-}354}$. The wild-type $K_{D(DNA)}$ was 0.67 μM, compared with Y89D $K_{D(DNA)}$ of 2.55 μM (Figure 5.6). These values are the concentrations at which 50% DNA substrate is bound, as calculated from the sigmoidal fit of DNA binding as shown in Figure 5.6.

5.6. Single Incorporation Kinetic Studies of PrimPol^{Y89D}

As a reduction in DNA binding is likely to affect the incorporation of nucleotides but is independent of catalytic turnover, we investigated the single-incorporation kinetics of this mutant variant of PrimPol. As previously shown in Section 4.3, we measured the single incorporation catalytic efficiency of wild-type PrimPol (k_{pol}) and its binding efficiency to incoming dNTPs ($K_{D(dNTP)}$). Wild-type PrimPol has a k_{pol} of 6.98 min^{-1} and binds the incoming dNTP with a $K_{D(dNTP)}$ of 15.51 μM. We determined the Y89D variant of PrimPol to be a slightly slower polymerase, with a k_{pol} of 4.07 min^{-1} (Figure 5.7) compared with wild-type 6.98 min^{-1}. However, this rate falls well within the same range of experimental error and therefore it is unlikely that the catalytic efficiency of Y89D is altered too much relative to the wild-type protein. In contrast, PrimPol^{Y89D} binds dNTPs over ten-fold less competently with a $K_{D(dNTP)}$ of 170.21 μM (Figure 5.7), compared to a

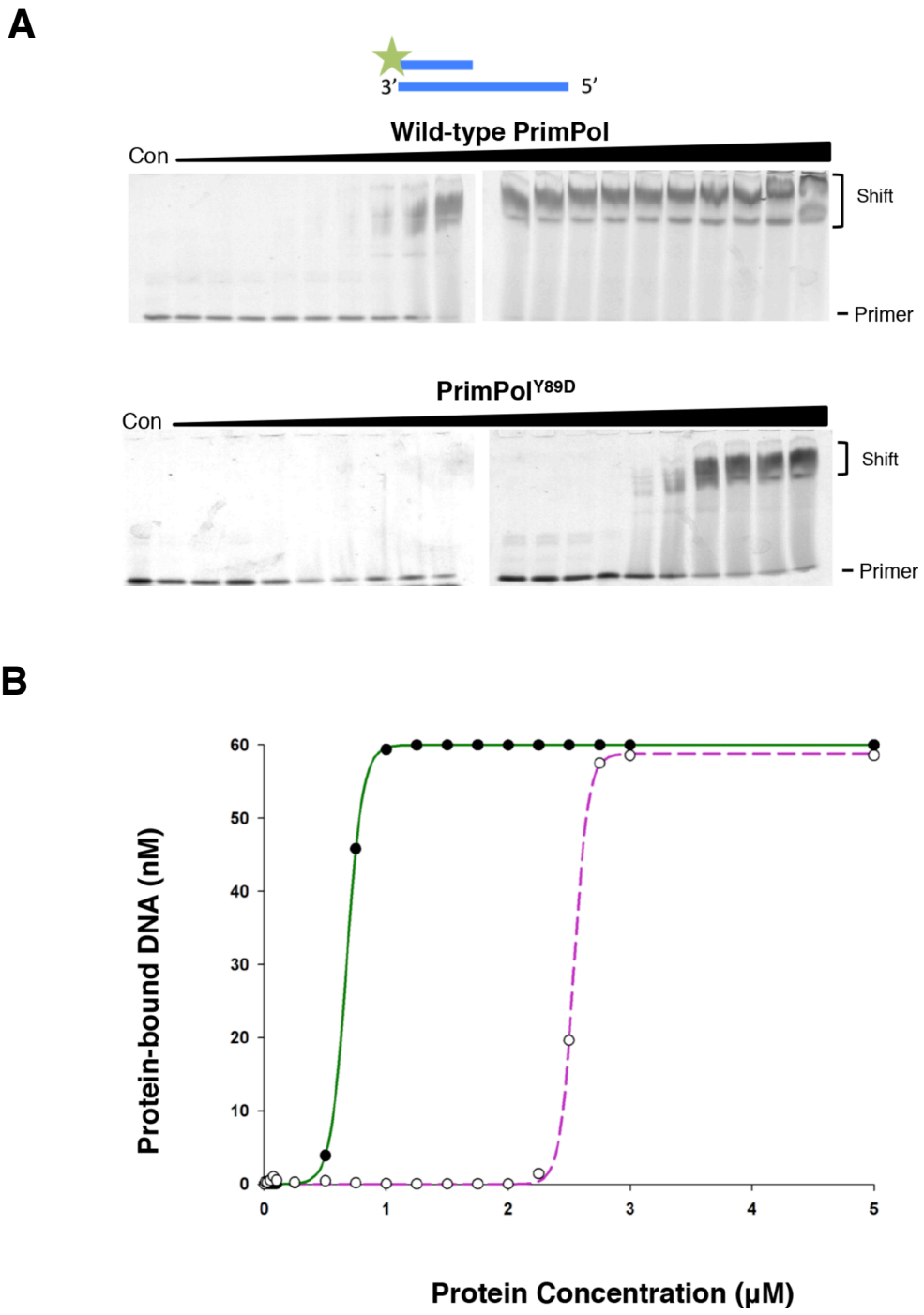


Figure 5.6. DNA binding efficiency of polymerase domains of wild-type PrimPol and PrimPol^{Y89D}.
(A) The polymerase domains (PrimPol$_{1–354}$) for both wild-type and PrimPol^{Y89D} were used to eliminate the effects of binding from the zinc finger domains. Wild-type PrimPol domain bound to DNA at a much lower concentration than PrimPolY89D as determined by EMSAs. The concentrations used in both of these EMSAs were 0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 5 and 10 μM. **(B)** These assays were subsequently quantified to determine $K_{D(DNA)}$ values for the polymerase domains of wild-type PrimPol (solid green line) and its Y89D variant (broken purple line). $K_{D(DNA)}$ of wild-type PrimPol was 0.69 μM, while the $K_{D(DNA)}$ of PrimPolY89D was 2.55 μM.

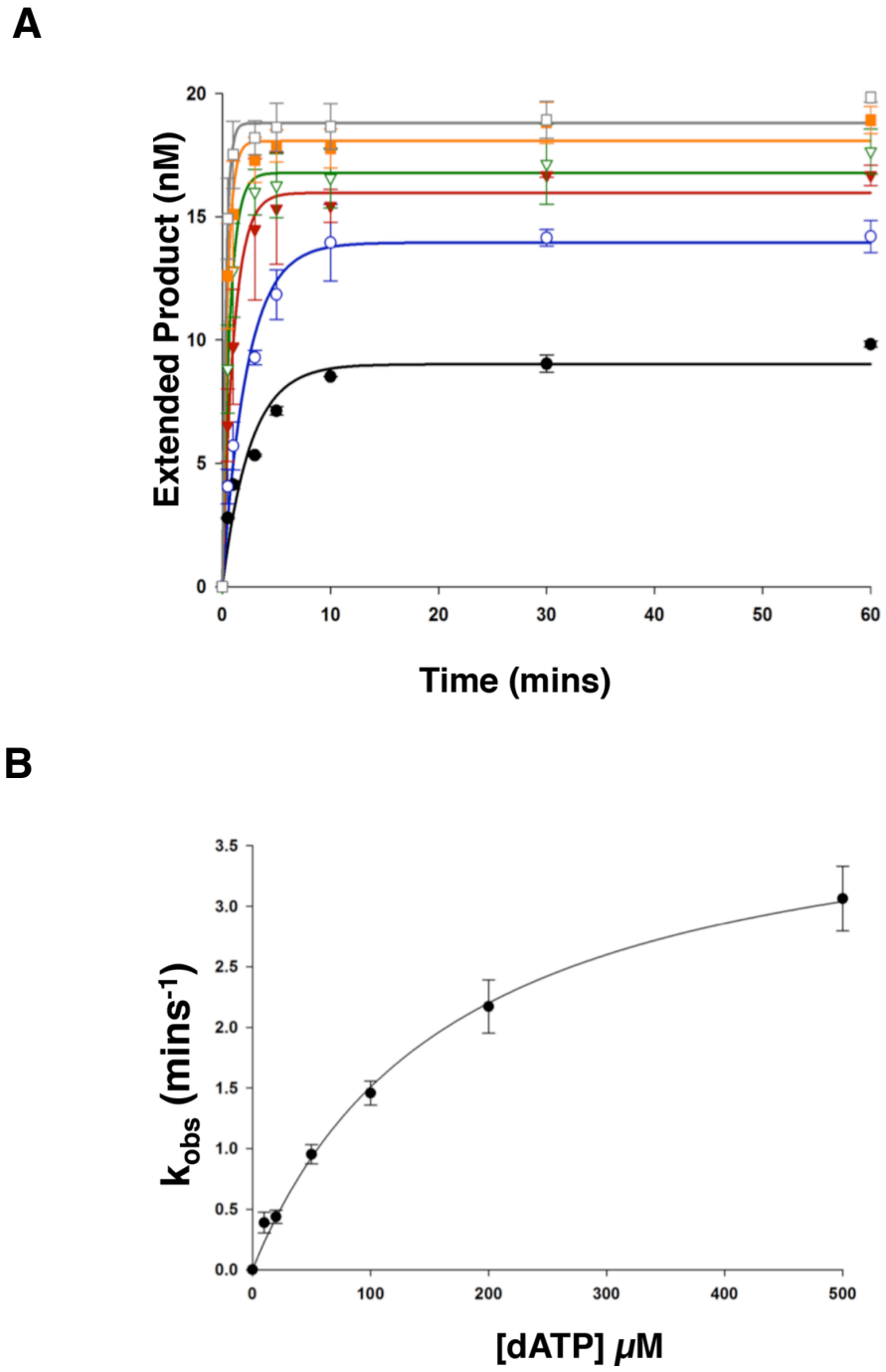


Figure 5.7. Single incorporation kinetics of PrimPol^{Y89D}.
Single turnover kinetics of PrimPol^{Y89D} were used to determine catalytic efficiency (k_{pol}) and dNTP binding efficiency ($K_{D(dNTP)}$) for comparison against the wild-type protein. (A) The concentration of extended DNA product over time, as determined by electrophoresis, was plotted against time and fit to an exponential curve for a number of dATP concentrations. These data were fit to exponential curves as described in the equation in Section 2.3.6. and k_{obs} was determined. Experiments were carried out in triplicate and error bars represent standard deviations. (B) k_{obs} was subsequently plot against dATP concentration for PrimPol and these data were fit to hyperbolic curves as described in Section 2.3.6. to determine k_{pol} and $K_{D(dNTP)}$. k_{pol} for PrimPol^{Y89D} was found to be 4.07 min^{-1} ± 0.19 and $K_{D(dNTP)}$ was 170.21 μM ± 18.80.

wild-type $K_{D(dNTP)}$ of 15.51 μM. Representative examples of the single incorporation experiments are shown in Appendix III. Further experiments are required to determine whether the dramatic reduction in dNTP binding may or may not be dependent on the initial binding of DNA.

5.7. PrimPol^{Y89D} Incorporation Fidelity Remains Unchanged

If dNTP coordination had been severely affected in the Y89D variant, we would expect quite a deviation from wild-type PrimPol in the fidelity of nucleotide incorporation. However, through single incorporation experiments to determine dNTP incorporation opposite different templating bases, we found that the Y89D variant of PrimPol did not have any appreciable variation from the wild-type enzyme in terms of fidelity (Figure 5.8), with PrimPol^{Y89D} not showing deviation from standard Watson-Crick base pairing.

5.7.1. PrimPol^{Y89D} Retains its Translesion DNA Synthesis Activities

If there were an alteration in the coordination of incoming dNTPs, there would also be a lower probability that this enzyme would be able to bypass the highly distortive lesions that the wild-type enzyme could bypass. As this enzyme had such low processivity, when the lesion bypass activity of PrimPol^{Y89D} was initially tested the primer extension activity was so low that the primers did not extend up to the lesion. As such, we used primers immediately 5' of the lesion to measure lesion bypass activity. We observed that PrimPol^{Y89D} was able to synthesize through even highly distortive lesions, such as a (6-4)PP (Figure 5.9), indicating that there is no defect in dNTP coordination.

5.7.2. PrimPol^{Y89D} Does Not Tolerate Mismatched Base Pairs

The primer DNA coordination could also be subject to a change in coordination relative to the incoming nucleotide for chain extension. DNA polymerases discriminate against mismatches both at the binding step and the insertion step (S. J. Johnson and Beese, 2004), so we would expect the Y89D variant of PrimPol to have a more significantly reduced catalytic efficiency than the wild-type protein than is evident if PrimPol^{Y89D} could tolerate mismatched base pairs. PrimPol^{Y89D} was assessed for its ability to extend from a range of templates that had a mismatched base at its primer-template junction, similar to the products you would create following a mismatched base incorporation, using different dNTPs. As expected, our data suggests that PrimPol^{Y89D}

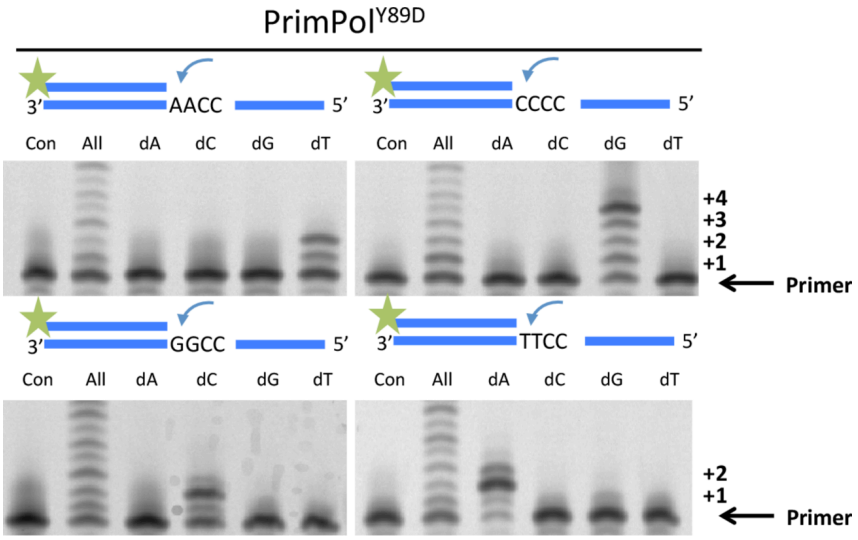


Figure 5.8. The fidelity of the PrimPol^{Y89D} variant remains unchanged relative to the wild-type PrimPol.
To provide comparable activity between the two constructs, the PrimPol^{Y89D} construct was incubated for 30 minutes along with individual dNTPs as indicated, all four dNTPs or a no dNTP control along with substrates 9-11, and 42, from Table 2.6 to allow incorporation of single nucleotides opposite two templating bases for each individual base. There was no deviation from the standard Watson-Crick base pairs from the PrimPol^{Y89D} variant.

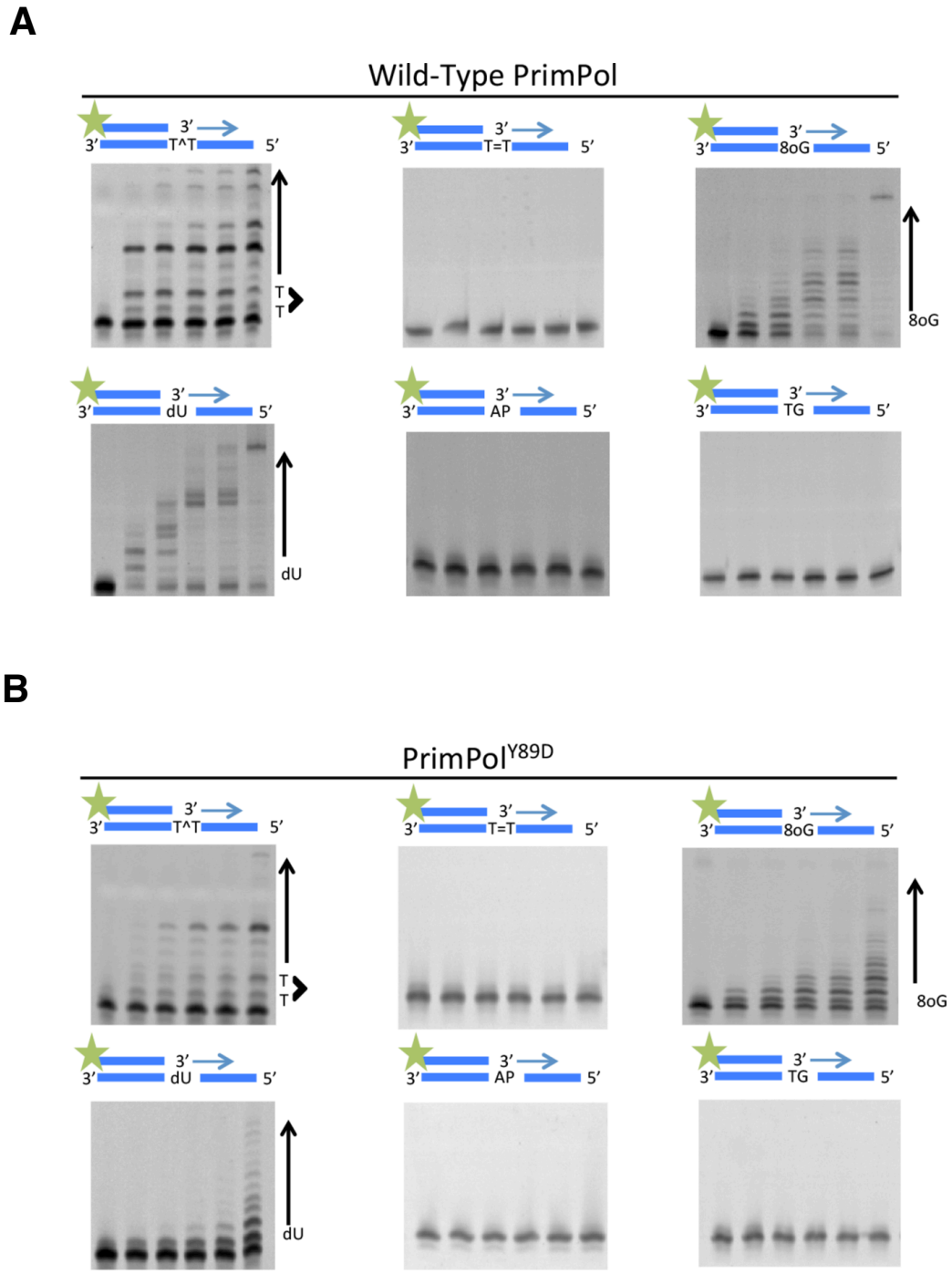


Figure 5.9. The translesion DNA synthesis spectrum of the PrimPol^{Y89D} variant remains unchanged.
To determine whether the PrimPol^{Y89D} variant retained its TLS activities, the substrates used required the lesions immediately downstream of the primer-template junctions; if not, the PrimPol^{Y89D} variant would not reach the lesion after 60 minutes. Substrates used were substrates 30, 32, 37, 40, 42 and 44 from Table 2.6. **(A)** PrimPol has the ability to bypass a 6-4 photoproduct (top-left), an 8-oxoguanine moiety (top-right) and deoxyuracil (bottom-left). Wild-type PrimPol cannot read-through a CPD (top-center), apurinic/apyrimidinic site (bottom-center) or a thymine glycol lesion (bottom-right). **(B)** PrimPol^{Y89D} retains its TLS activity as the wild-type protein, albeit with lower polymerase activity. In each case, the first lane represents a control and lanes 2-6 represent incubation periods of 0.5, 1, 3, 5 and 60 minutes respectively.

has little tolerance for mismatched base pairs (Figure 5.10), no different to the wild-type enzyme.

5.7.3. PrimPol^{Y89D} Cannot Utilise rNTPs for Primer Extension

Aromatic residues sometimes act as a steric gate for sugar discrimination (Brown and Suo, 2011). Mutation of these key residues can result in the incorporation of rNTPs instead of dNTPs opposite a templating DNA strand (Donigan et al., 2014). PrimPol^{Y89D}, like wild-type PrimPol, was also unable to extend primers using rNTPs (Figure 5.11). Therefore, it is highly unlikely that this tyrosine residue is involved in discerning the sugar of the incoming nucleotide, particularly as the Y89D variant also loses its ability to prime using rNTPs, as illustrated in Section 5.3.

5.8. The PrimPol^{Y89D} Mutation Induces a Significant Structural Alteration in PrimPol

The observed reduced enzyme activity of PrimPol^{Y89D} could be the result of the loss of aromatic interactions involved in local base stacking events with either the template or the incoming nucleotide. However, it may instead be required for maintaining the structural integrity of the enzyme and mutation of this tyrosine causes more global effects in the structure of PrimPol. To test this second possibility, we subjected the polymerase domains of both the wild-type and PrimPol^{Y89D} to thermal denaturation and circular dichroism (CD) to check for alterations. There was a striking difference in both the thermal denaturation profiles and the CD spectra of the two enzymes (Figure 5.12A), suggesting this point mutant results in global changes to the enzyme that causes the reduction in activity observed above. The melting point of the polymerase dropped from 49.12°C for the wild-type polymerase domain compared to 42.50°C for PrimPol^{Y89D} (Figure 5.12B). This significant drop in melting temperature suggests that the enzyme has become less stable and more structurally disordered following the point mutation. Analyses of the secondary structure from the circular dichroism spectra of the two variants' polymerase domains by the variable selection algorithm (CDSSTR) (Lobley et al., 2002), which provides better fits for globular proteins, revealed a difference in the α-helical content of these domains (37% for wild-type PrimPol compared with 21% for Y89D), as well as β-strand content (25% for wild-type compared with 29% for Y89D). Additionally, this analysis predicted 30% of the structure

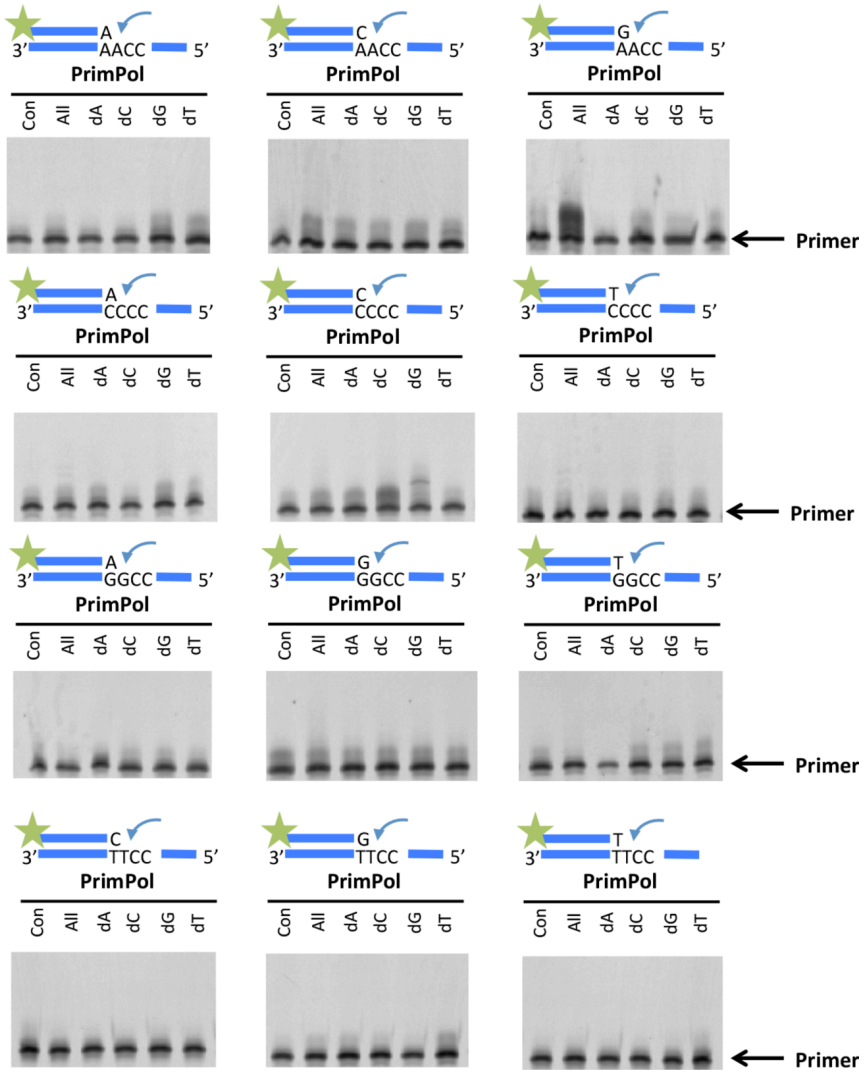


Figure 5.10. PrimPol^{Y89D} does not exhibit any mismatch tolerance activities. PrimPol^{Y89D} was incubated for 30 minutes with individual dNTPs as indicated, all four dNTPs or a no dNTP control along with the mismatched substrates from substrates 13-28 from Table 2.6. These substrates represent primers annealed to templates with the primer template junction representing each of the 12 possible combinations of base pairs that exclude the 4 Watson-Crick base pairs. PrimPol^{Y89D} does not extend any of the substrate primers with any proficiency.

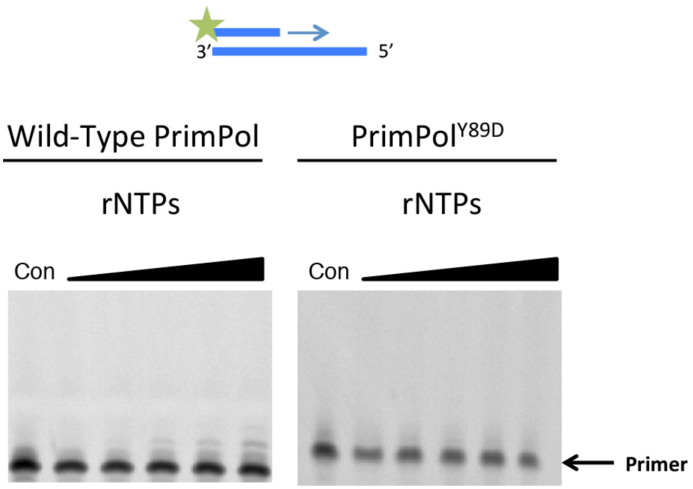


Figure 5.11. PrimPol^{Y89D} does not incorporate ribonucleotides.
PrimPol^{Y89D} was assessed for its DNA-dependent rNTP incorporation activity. PrimPol, and PrimPol^{Y89D} were incubated with rNTPs and an undamaged DNA substrate (substrate 1 from Table 2.6) at 0.5, 1, 3, 5 and 60 minute time points alongside a control containing no protein. PrimPol^{Y89D}, like wild-type PrimPol, is unable to extend a DNA primer using ribonucleotides opposite a DNA template.

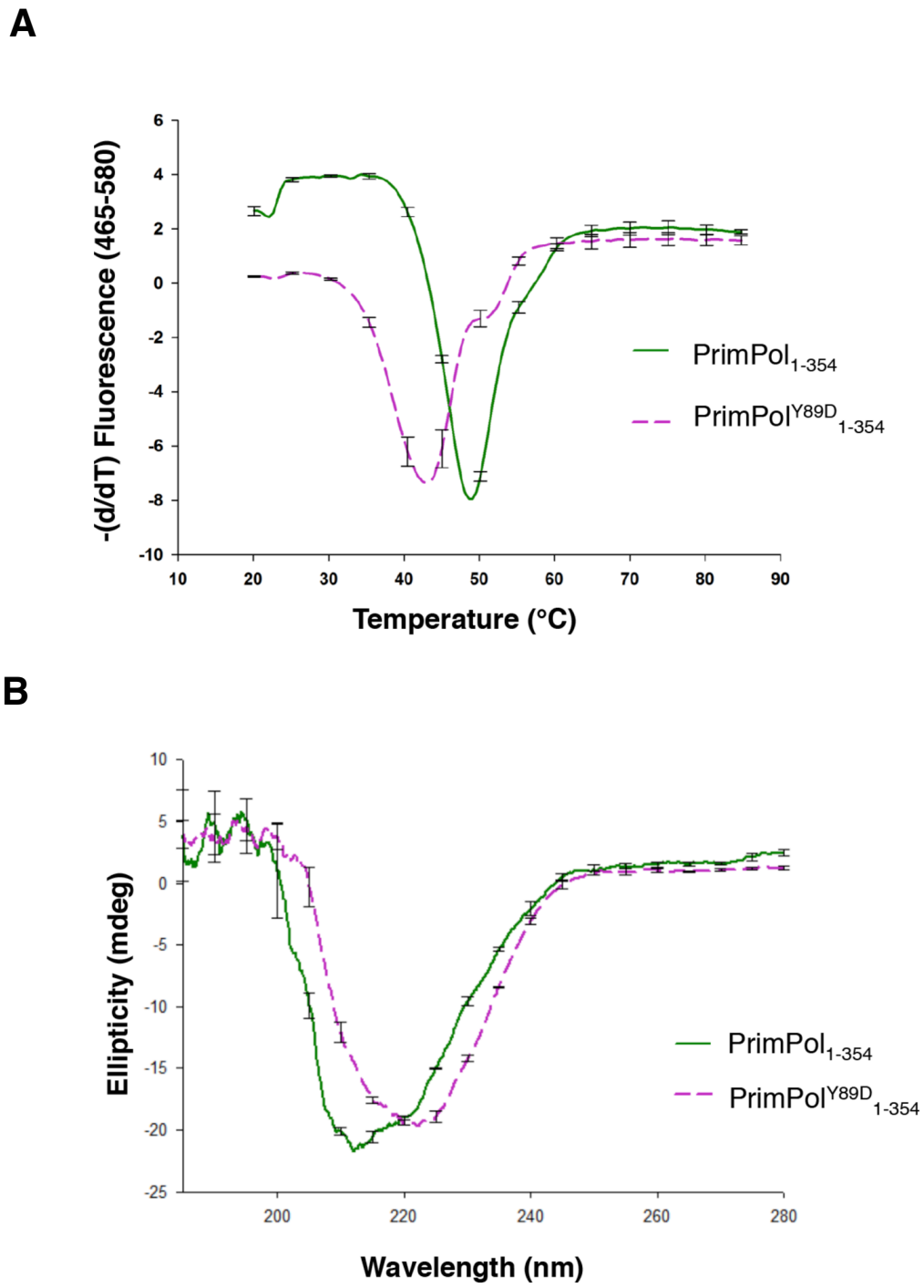


Figure 5.12. Y89D induces a structural perturbation in PrimPol.
The polymerase domain of the Y89D variant of PrimPol has a different global structure and thermal stability compared to wild-type PrimPol. **(A)** Thermal denaturation of the polymerase domain of PrimPol and its Y89D variant. The negative differential of the fluorescence profiles of the polymerase domains ($PrimPol_{1–354}$) of wild-type PrimPol (solid green line) and its Y89D (broken purple line) variant plotted against temperature to determine their melting points. The melting point of the polymerase domain of the wild- type enzyme was found to be 49.12◦C, while the Y89D variant had a considerably lower melting point at 42.50◦ C. **(B)** Circular dichroism spectra of the polymerase domains of wild-type PrimPol and $PrimPol^{Y89D}$ also show a substantial difference in the global structure of $PrimPol^{Y89D}$ relative to the wild-type enzyme.

of the PrimPol^{Y89D} polymerase domain to be unstructured, whereas just 24% of wild-type was predicted to be unstructured.

5.9. PrimPol^{Y89D} Significantly Decreases DNA Replication Fork Rates *in vivo*

To gain further insights into how the Y89D mutant form of PrimPol may contribute to the myopic phenotype observed in human patients, we studied the influence of PrimPol^{Y89D} on DNA replication in damaged and undamaged cells. We utilised the DT40 PrimPol deleted cell line (PrimPol $^{-/-}$) (Bianchi et al., 2013) to examine the ability of PrimPol^{Y89D} to complement these cells, as previously reported for other mutant forms of PrimPol (Sections 3.11 and 4.10). These studies were carried out in collaboration with Dr. Laura Bailey from our laboratory. We stably expressed PrimPol^{Y89D} at similar levels to wild-type PrimPol in PrimPol knock-out DT40 cells (Figure 5.13A), with no obvious effects on cell growth.

We reported previously that PrimPol$^{-/-}$ DT40 cells have a greater sensitivity to UV irradiation (Section 4.10) thus we first examined the effects of increasing doses of UV-C on PrimPol^{Y89D} complemented cell lines. Although PrimPol$^{-/-}$ cells complemented with wild-type PrimPol showed a reversal of this sensitivity, cells complemented with PrimPol^{Y89D} maintained a decreased viability after UV damage (Figure 5.13B). Next, we measured the replication fork rates in PrimPol complemented cell lines using DNA fibre analysis. Strikingly, we observed a dramatic decrease in fork speeds in cells complemented with PrimPol^{Y89D} (average 0.84 kb/min) compared with cells carrying the wild-type PrimPol (average 1.66 kb/min) (Figure 5.13C). This decrease in fork speed was significantly greater than that previously reported in PrimPol$^{-/-}$ cells (average 1.45 kb/min) (Bianchi et al., 2013), suggesting the Y89D mutant form of PrimPol may actually impede on-going DNA replication. In contrast, PrimPol^{Y89D} complemented the increased stalling defect observed in PrimPol$^{-/-}$ cells treated with UV-C, equally as well as wild-type PrimPol (Figure 5.13D), demonstrating that the polymerase activity, rather than the primase activity, is much more considerably affected by the Y89D mutation.

5.10. Summary and Discussion

A recent exome sequencing study has implicated a point mutation (Y89D) in the DNA damage tolerance polymerase, PrimPol, in high myopia in human patients (Zhao et al.,

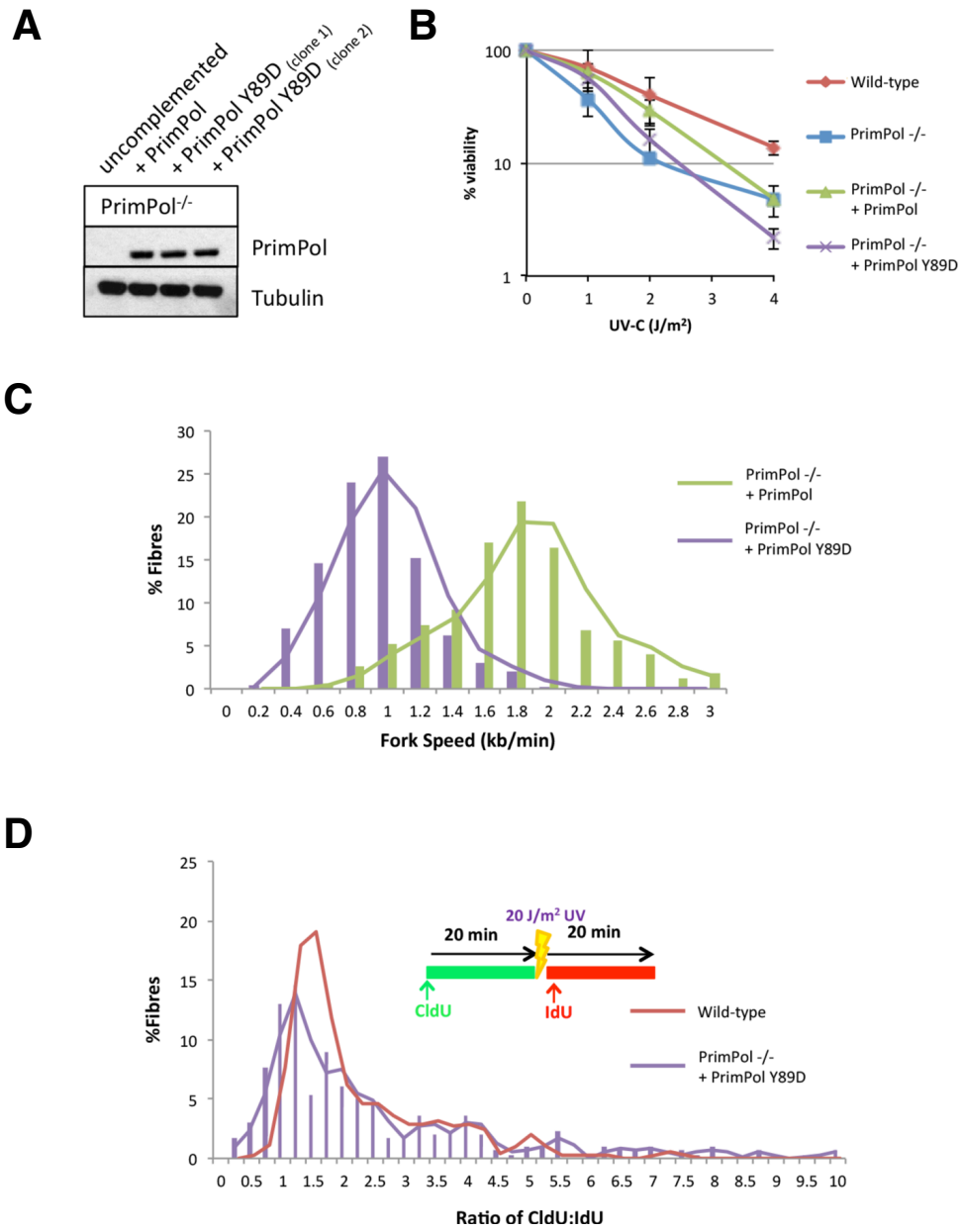


Figure 5.13. PrimPol^{Y89D} cannot fully complement the UV sensitivity seen in PrimPol$^{-/-}$ DT40 cells due to decreased replication fork speeds.
(A) Western blot showing overexpression of wild type and Y89D point mutant forms of PrimPol in PrimPol knockout (PrimPol$^{-/-}$) DT40 chicken cells. Primary antibodies used are anti-PrimPol and anti-tubulin. **(B)** Cell viability was measured using CellTiter-Blue 48 hours after increasing doses of UV-C were administered to the different cell types. Results represent three independent repeats and the error bars represent the standard deviation across the repeats. **(C)** DNA fibres were measured after sequential labelling with CldU and IdU for 20 min each to calculate replication speed. Results represent three independent repeats. There is a severe hindrance in unperturbed replication speed. **(D)** A 20 J/m^2 UV-C pulse was included between the two labels and the ratio of the pre- and post-UV-C labels was calculated, with increased ratios representing replication stalling after UV. Results represent three independent repeats. Experiments shown here were carried out by Dr. Laura Bailey and fibre length analysis was carried out in collaboration by Dr. Laura Bailey and Benjamin Keen.

2013). However, this report did not investigate the molecular consequences of this genetic alteration. In this chapter, we report that the Y89D variant of PrimPol, previously found to be associated with high myopia, is a low processivity variant of PrimPol. Unlike wild-type PrimPol, it inserts just one nucleotide in a single binding event and has a significantly lower affinity for both dNTPs and DNA. Notably, this mutant did not exhibit any alteration in the fidelity, mismatch tolerance or translesion DNA synthesis spectrum relative to wild-type PrimPol. To determine whether this altered activity was a result of losing the hydrophilic hydroxyl moiety or the hydrophobic ring moiety, we constructed two more variants in which tyrosine 89 was mutated to either phenylalanine or serine. The Y89F mutant restored processivity to near-wild-type levels but Y89S had a similar activity to Y89D. In addition, we found that the Y89D PrimPol variant was unable to produce ribonucleotide primers but this activity was restored in the Y89F variant, indicating that PrimPol requires a hydrophobic aromatic ring at position 89 to maintain processivity but the presence of a hydroxyl moiety is dispensable.

Insights from structural studies on primases and polymerases suggest two different potential explanations for the reduction in polymerase processivity observed in this study. The conserved tyrosine at position 89 could have an important role in base stacking or could be required for the structural integrity of the protein. There is an invariant phenylalanine in NHEJ AEPs that plays a vital role in the recognition and orientation of both the templating and the incoming bases through stacking interactions with both bases and is responsible for moving the incoming nucleotide into the enzyme's active site (Brissett et al., 2011; Pitcher et al., 2007). Phenylalanine 290 in human Pol η helps form the hydrophobic core of the thumb domain (Biertümpfel et al., 2010). Mutation of this residue is associated with the genetic disorder *xeroderma pigmentosum variant* (XPV), which is characterised by a susceptibility to skin cancers caused by a failure to perform proficient TLS across UV lesions. This mutation negates polymerase activity and is predicted to destabilise the thumb domain, resulting in reduced primer binding. Whilst the structure of PrimPol has not yet been elucidated, it is conceivable that the tyrosine at position 89 is involved in preserving the structure of a protein region analogous to the thumb domain that is found in other polymerases involved in binding the templating DNA strand. Tyrosine 89 is situated in a conserved hydrophobic pocket containing other conserved aromatic residues including Y87, F88 and Y90, as well as a hydrophobic residue at position 86 (Figure 5.1) and it is likely that

disruption of Y89 disrupts a conserved interaction between these amino acids. A secondary structure prediction from PSIPRED (Jones, 1999) predicts that all of these amino acids lie within an α-helical region and the Y89 mutation may disrupt this helix. The circular dichroism data certainly supports this model, indicating that the Y89D variant has less α-helical structure than the wild-type. We postulate that the Y89D mutation causes a disruption in the global structure of the polymerase domain of PrimPol. As the affinity for DNA and dNTPs is decreased in the Y89D variant, but the apparent fidelity of the enzyme remains unchanged, we suggest that dNTP binding is dependent on DNA binding, which is disrupted by the structural changes induced by this mutation.

In vivo analysis identified a dramatic decrease in replication fork rates in cells expressing PrimPol^{Y89D} and, despite being able to complement fork stalling in PrimPol$^{-/-}$ cells after UV damage, it was unable to prevent decreased cell viability. Consistent with *in vitro* assays, PrimPol^{Y89D} causes decreased replication fork rates in cells and it appears that this defect is enough to decrease the damage tolerance of these cells. Both our *in vivo* and *in vitro* studies on this mutant indicate that although PrimPol^{Y89D} retains the ability to bypass lesions, its capacity to replicate DNA is severely compromised. Therefore, it is appears that although PrimPol^{Y89D} is recruited to stalled replisomes, the underlying replication defect causes a significant slowing of replication forks, potentially resulting in incomplete lesion bypass, causing cells to progress through the cell cycle with under replicated genomic DNA. Clearly, these cells carry an abundance of PrimPol compared to wild-type levels and thus replication fork speed alone in endogenous mutants may not be so significantly altered. However, in cases where PrimPol is actively recruited to on-going forks or damage sites, this ‘slow’ mutant is likely to cause significant problems. Thus, unexpectedly, it appears that harbouring the PrimPol^{Y89D} mutation is more detrimental than not having PrimPol at all. PrimPol is not present in cells in particularly high abundance and is likely to be tightly regulated in S/G2 phase; an artificial overexpression of this mutation in these experiments is likely to contribute to the defects observed here. This data correlates with the phenotypes observed in patients carrying this mutation as problems are observed in tissues where levels of UV damage is expected to be high. Thus, in tissues where PrimPol is recruited in higher abundance to replicate through DNA damage, such as eyes, cellular defects are much more likely to be prevalent. In accordance, the previous study finds

that the high myopia phenotype is late onset (Zhao et al., 2013), which is concurrent with the accumulation of DNA damage over time.

The link between reduced PrimPol polymerase activity and myopigenicity remains unclear. Another ocular disorder has been linked to mutations in a polymerase gene, progressive external ophthalmoplegia (PEO) is linked to mutations in the gene for pol γ in its catalytic domain (Van Goethem et al., 2001). Mutation results in an autosomal dominant phenotype and is characterised by deletions in mtDNA copy number (Van Goethem et al., 2001). Similar to the PrimPol variant phenotype this disorder is late onset, typically appearing in adults above the age of 18 with patients progressively deteriorating over time. However, it is likely that the two disorders are unrelated both physiologically and at a molecular level. The high myopia is not related to muscle deterioration whereas PEO patients develop a weakness in the muscles involved in eye movement and generally also exhibit skeletal muscle weakness that is particularly noticeable during exercise due to their mitochondrial defects. Additionally, the major mutation in pol γ that cause PEO are the result of enhanced incoming nucleotide binding, which results in a 10-100 fold drop in polymerase fidelity (Chan and Copeland, 2009), a characteristic that is unchanged in PrimPol^{Y89D}. However, more severe forms of PEO are caused by mutations that result in a reduction in pol γ activity, similar to the reduction in activity evident in PrimPol^{Y89D} and whilst we have the attributed the physiological effects of PrimPol^{Y89D} to a reduction in fork rate, it would be interesting to see what effect, if any, PrimPol^{Y89D} has on mitochondrial copy number.

Previous studies into the genes that predispose high myopia have identified two clusters of mutations, one in the outer retina affecting photoreceptors and ON-bipolar cell function and another in the scleral extracellular matrix composition and metabolism (Morgan et al., 2012). Significantly, mutations in PrimPol do not fit into either category. The severity of the clinical phenotype of patients with the Y89D mutation varied (Zhao et al., 2013) and it is highly likely that this gene is not a singular, unique aetiological factor and that other genetic and environmental factors also contribute to pathogenesis; however, it is evident from this previous study that the singular homozygote had a more severe phenotype than the heterozygotes that were identified. Although replication defects associated with PrimPol^{Y89D} suggests that it contributes to the reported clinical phenotypes, further work is required to establish if a direct link exists between replication stress and onset of high myopia.

CHAPTER 6

Characterisation of the Molecular Interactions Between PrimPol and Replication Protein A

6.1. Introduction

Translesion DNA synthesis polymerases are low fidelity enzymes that require strict regulation to reduce their mutagenic impact on the genome. The predominant mechanism for the regulation of polymerases is through the monoubiquitination of PCNA (Section 1.6) (Lehmann et al., 2007); when the replication fork reaches sites of DNA damage PCNA becomes monoubiquitinated and this allows access to the distributive TLS polymerases to synthesise DNA across the site of damage before the replication fork can continue to replicate the genome in a processive manner. PrimPol is a novel primase-polymerase involved in DNA damage tolerance through both TLS and post-lesion priming mechanisms (Chapter 4). In this chapter the means through which PrimPol is regulated is examined.

Although PrimPol has been shown to co-localise with PCNA upon DNA damage (Mourón et al., 2013), there remains no evidence that PrimPol interacts with PCNA, monoubiquitinated or otherwise. It has recently been shown that, unlike other TLS polymerases, PrimPol is not stimulated by monoubiquitinated PCNA and does not interact with PCNA *in vivo* (Guilliam et al., 2015). To identify potential interaction partners, Dr. Sean Rudd from our laboratory carried out a mass spectrometry analysis of the proteins that were pulled down from human kidney cells (HEK 293) stably overexpressing Strep-tagged PrimPol and identified components of the replication protein A (RPA) complex as potential binding partners for PrimPol (Guilliam et al., 2015; Rudd, 2013). Concurrently, a separate mass spectrometry analysis of an RPA pull-down from HEK 293T cells identified PrimPol as an interacting partner and that this interaction is required for foci formation *in vivo* (Wan et al., 2013). RPA has previously been shown to stimulate the activities of other polymerases, such as pols α and δ (Braun et al., 1997; Kenny et al., 1989; Tsurimoto and Stillman, 1989).

The most abundant single-stranded binding protein in the cell (Seroussi and Lavi, 1993), RPA is heterotrimeric complex composed of subunits 70, 32 and 14 kDa in size (Fairman and Stillman, 1988), henceforth referred to as RPA70, RPA32 and RPA14 respectively. RPA is essential for a large number of genomic DNA metabolic reactions, including replication, recombination and DNA repair mechanisms NER, BER, MMR and double strand break repair (Reviewed in (Zou et al., 2006)). It is responsible for coating ssDNA in response to the generation of DNA lesions to prevent aberrant DNA structure

formation and to recruit DNA damage response proteins (Majka et al., 2006; Zou et al., 2006).

The RPA heterotrimer is comprised of 6 oligosaccharide/oligonucleotide-binding (OB) folds, four in RPA70 (RPA70N, RPA70A, RPA70B and RPA70C) and one in each of RPA32 and RPA14 (Figure 6.1). The OB folds are five β sheets oriented in a β-barrel arrangement, a common structural arrangement amongst oligonucleotide binding proteins (Bochkarev and Bochkareva, 2004; Bochkareva et al., 2002). RPA70A and RPA70B are responsible for the initial binding to single-stranded DNA, before switching to a more stable DNA binding mode involving the RPA70C and RPA32 OB folds (Bochkareva et al., 2002). RPA14 is important for trimerisation though its other functions remain unclear. RPA70N is involved in protein-protein interactions and mediates the interactions with a number of DNA damage response proteins, including p53, ATRIP, RAD9 and MRE11 (Feldkamp et al., 2013; Xu et al., 2008). In addition to its OB folds, RPA also possesses a winged-helix domain at the C-terminus of RPA32. This domain is also involved in protein-protein interactions with a number of DNA repair proteins including Uracil-DNA glycosylase 2, XPA, and RAD52 (Mer et al., 2000), all of which operate in different repair pathways.

In this chapter, we define the binding site of PrimPol on its replication protein A binding partner and identify two protein-interaction motifs in the C-terminus of PrimPol that mediate the formation of this complex. The data reported here acts to complement the findings by others in our group and our collaborators that were recently published (Guilliam et al., 2015). At the end of this chapter, two potential models of regulation of the activities of PrimPol are discussed based on the work presented in this, and previous chapters.

6.2. The C-terminus of PrimPol is Poorly Structured

As the C-terminus of PrimPol, downstream of the zinc finger (amino acids 488-560), is not predicted to fold into a structural or catalytic domain, based on bioinformatics analysis of its primary sequence alone, we decided to determne whether the C-terminal portion of the protein acted to reduce the overall stability of the protein. Thermal denaturation studies have shown that there is a considerable increase in the stability of PrimPol upon removal of amino acids 488–560 (Figure 6.2), implying that this region lacks discernible structure that would contribute to protein stability. Disordered regions

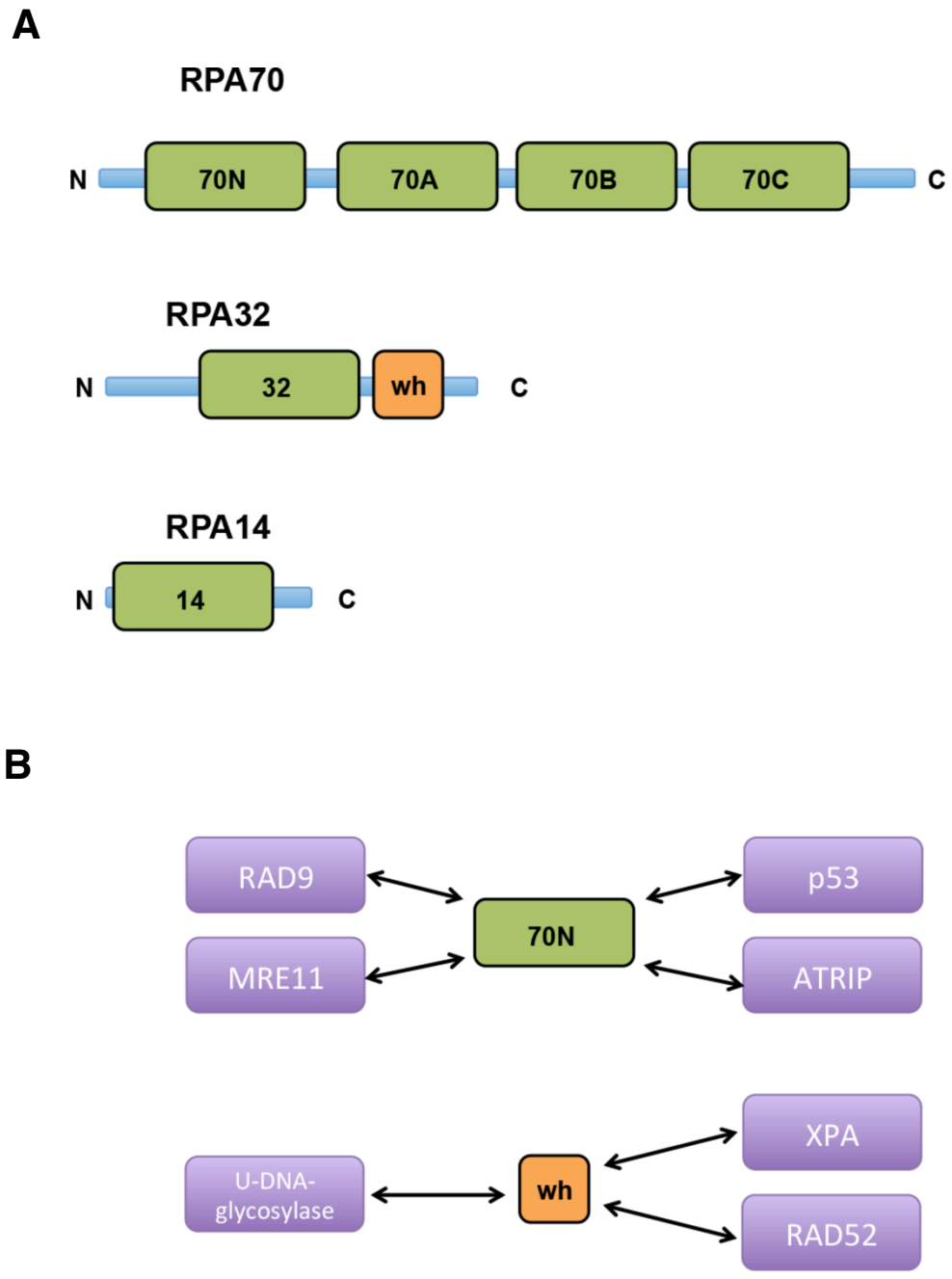


Figure 6.1. The domain architecture and protein interaction domains of RPA. **(A)** RPA is a heterotrimeric single-stranded DNA binding protein, composed of three domains of 70 kDa (RPA70), 32 kDa (RPA32) and 14 kDa (RPA14). RPA70 consists of four OB folds, indicated in green, RPA70N, RPA70A, RPA70B and RPA70C. RPA32 consists of a single OB fold, again indicated in green, and a C-terminal winged-helix domain (orange). RPA14 consists of a single OB fold, shown in green. **(B)** The RPA70N OB fold and the winged-helix domain of RPA32 are primarily responsible for RPA's interaction with other proteins. RPA70N has been shown to interact with RAD9, MRE11, p53 and ATRIP. The winged-helix domain has been shown to interact with uracil-DNA-glycosylase, XPA and RAD52.

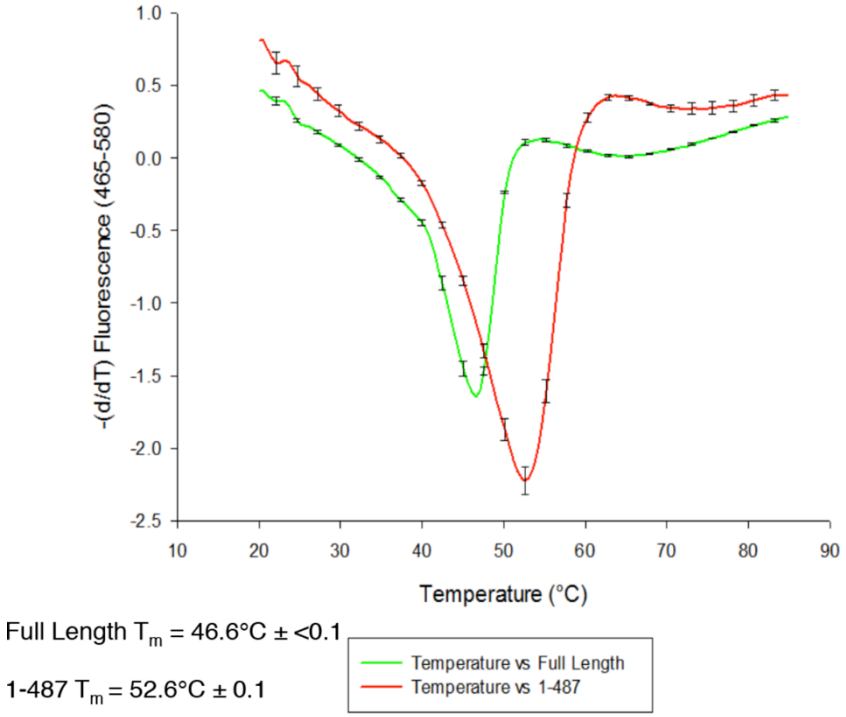


Figure 6.2. The C-terminus of PrimPol is an unstructured, low-stability region of protein.
Wild-type PrimPol and PrimPol$_{1\text{-}487}$ were analysed by thermal denaturation. Full length PrimPol has a significantly lower melting temperature (46.6°C ± <0.1) than PrimPol$_{1\text{-}487}$ (52.6°C ± 0.1), which suggests that the amino acids 487-560 at the C-terminus do not form a secondary structure and are instead unstructured regions of protein.

of proteins often serve as scaffolds involved in protein–protein interactions and we speculate that this portion of human PrimPol might also act as an interaction interface. Indeed this is in agreement from the work from Wan et al., who found that the final 80 amino acids of PrimPol (480-560) are sufficient for RPA-dependent foci formation following DNA damage (Wan et al., 2013).

6.3. Identification of PrimPol and RPA Reciprocal Binding Domains

The mass spectrometry data from both our group, and also Wan *et al.*, revealed that human PrimPol interacts with the RPA70 subunit of RPA (Guilliam et al., 2015; Wan et al., 2013). Wan *et al.* reported that RPA70 lacking its RPA70C domain failed to co-immunoprecipitate with PrimPol, suggesting that this is the domain that interacts with PrimPol (Wan et al., 2013). However, in the same study, following hydroxyurea treatment in the presence of this RPA70 mutant lacking the RPA70C domain, PrimPol was still able to form nuclear foci, an apparent contradiction to their conclusion that it is the RPA70C domain that interacts with PrimPol (Wan et al., 2013). Furthermore, there have been no previous reports that RPA70C acts as a protein-protein interaction domain.

6.3.1. PrimPol Binds to RPA70N

To identify the reciprocal binding domains of human PrimPol and RPA, we established a collaboration with the Chazin group (Vanderbilt University, Nashville, USA) who employ nuclear magnetic resonance (NMR) approaches to characterise RPA interactions. Aaron Ehlinger in the Chazin group performed the NMR-based binding studies. As discussed above, the major protein-protein interaction domains of RPA are the OB fold RPA70N and the winged helix domain of RPA32, and the NMR studies were focused on these two domains. ^{15}N-^{1}H Heteronuclear Single Quantum Coherence (HSQC) NMR spectra of ^{15}N-enriched RPA70N (RPA70 1-120) and the winged helix domain of RPA32 (172-270) were acquired in the absence and presence of 2-fold molar excess of unlabelled PrimPol. Amide chemical shifts are sensitive to their local chemical environment and can be measured by NMR spectroscopy. When a ligand is bound, the location and/or intensity of NMR peaks at the binding site are altered. In addition, chemical shift perturbations can also be induced by the structural changes in the protein that are often brought about during ligand binding.

The NMR analysis of $RPA32_{172-270}$ did not reveal any significant amide chemical shifts between the absence and presence of PrimPol (Figure 6.3A), chemical shifts were of comparable location and intensity, demonstrating no perturbation by the presence of PrimPol. However, the NMR analysis of RPA70N following the addition of PrimPol showed a significant difference in the amide chemical shifts, specifically in an attenuation of peak intensity (Figure 6.3B). This decrease in peak intensity corresponds to a reduction in protein tumbling when part of a larger complex with PrimPol. PrimPol therefore interacts with the RPA70N OB fold of RPA as opposed to the RPA32 winged helix domain of RPA.

6.3.2. RPA70N Binds the C-terminus of PrimPol

HSQC NMR spectra were also measured on ^{15}N-enriched RPA70N in the absence and presence of our purified $PrimPol_{1-487}$ recombinant protein (Figure 6.3C; see Section 3.3.1 for purification of $PrimPol_{1-487}$). The spectrum in the presence of $PrimPol_{1-487}$ closely resembles that of RPA70N alone, demonstrating that the C-terminal portion of PrimPol is indeed required for RPA binding and that binding is abolished in its absence. Taken together, these results indicate that the unstructured C-terminal portion of PrimPol and the RPA70N OB fold of RPA mediate the interaction between PrimPol and RPA.

6.3.3. Purification of $PrimPol_{480-560}$

As these results from the Chazin group indicated that the final 80 amino acids of PrimPol are the residues responsible for RPA70N binding, we next aimed to study formation of this complex more directly. We constructed and purified $PrimPol_{480-560}$ as described in Tables 2.3-2.5 (Figure 6.4A). The protein has a predicted size of 10,997 Da but the protein bands run at a size closer to ~20kDa during SDS-PAGE analysis. To confirm this protein was indeed the correct protein, in addition to confirmation of the plasmid DNA sequence, we carried out two western blots on this protein, one to confirm the presence of PrimPol and another to confirm the presence of the 6x-His tag used to purify the protein (Figure 6.4B). In addition to the $PrimPol_{480-560}$ protein, the Chazin group provided us with RPA70N protein for our studies (Figure 6.4C).

6.3.4. PrimPol 480-560 Interacts with RPA70N

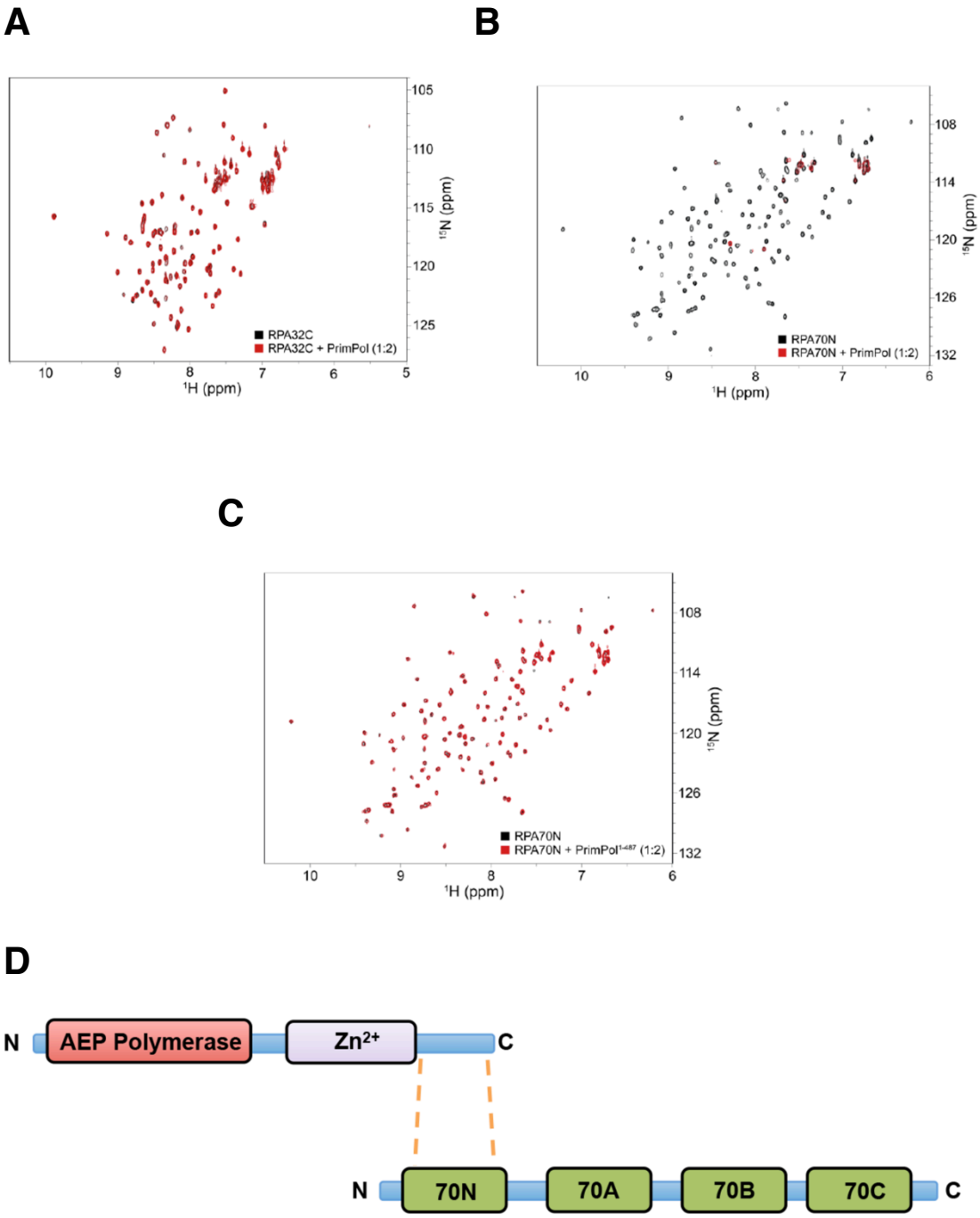


Figure 6.3. The C-terminus of PrimPol interacts with RPA70N.
(A) HSQC NMR spectra of ^{15}N-enriched winged helix domain of RPA32 (denoted RPA32C above) in the absence (black) and presence (red) of PrimPol in a 2:1 ratio. There are no significant changes in the NMR spectra upon titration of PrimPol. **(B)** HSQC NMR spectra of ^{15}N-enriched N-terminal OB fold of RPA70 (RPA70N) in the absence (black) and presence (red) of PrimPol. Upon titration of PrimPol there is a significant signal attenuation, suggesting an interaction between PrimPol and RPA70N. **(C)** This signal attenuation is not observed when PrimPol$_{1\text{-}487}$ is titrated in (red) rather than wild-type PrimPol. **(D)** The C-terminal region downstream of the zinc finger is essential for interaction with RPA70N. PrimPol does not interact with RPA32. The experiments shown above were carried out by Dr. Aaron Ehlinger from the Chazin laboratory.

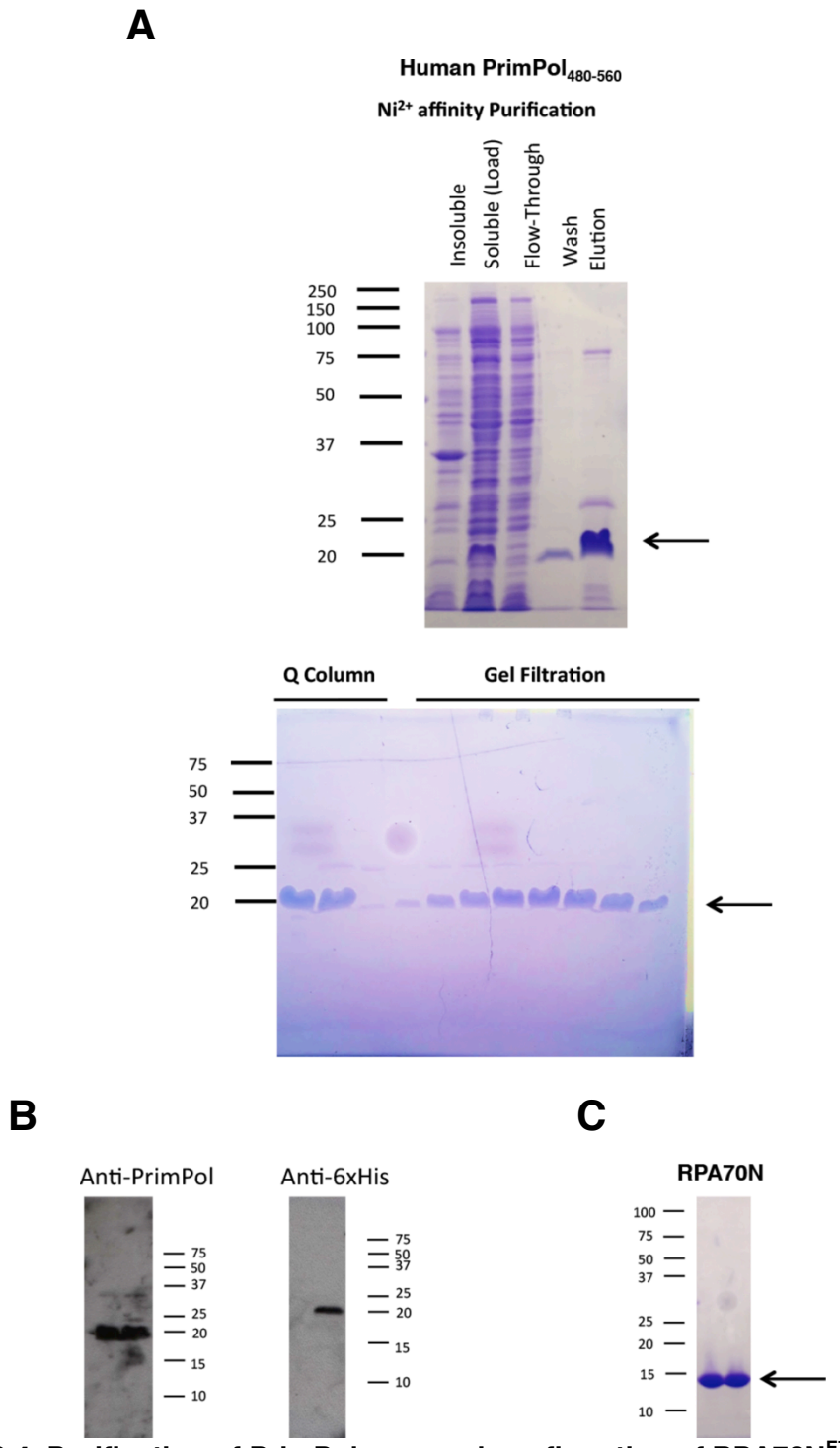


Figure 6.4. Purification of PrimPol$_{480-560}$ and confirmation of RPA70N^{E7R} purity. **(A)** PrimPol$_{480-560}$ (11.0 kDa) was purified by IMAC (Ni^{2+}) affinity, anion exchange and size exclusion chromatography. SDS-PAGE analysis of samples from purification fractions indicate a protein of size ~20 kDa. **(B)** The protein was confirmed correct by western blotting analysis through the use of a polyclonal Anti-PrimPol antibody and an antibody against the hexahistidine tag used to purify the protein. DNA sequence analysis of the expression vector and mass spectrometry were also utilised to confirm that this was indeed the correct protein. **(C)** SDS-PAGE analysis of purified RPA70N. RPA70N^{E7R} (13.2 kDa) was gifted by the Chazin laboratory and analysed by SDS-PAGE for purity and for any breakdown in transit.

To directly assess the interactions between these proteins, we utilised an analytical gel filtration chromatography approach. An S75 gel filtration column was initially calibrated through the size exclusion fractionation of a number of protein standards, albumin, ovalbumin, chymotripsinogen A, ribonuclease A and aprotinin (Figure 6.5A). The presence of these proteins in the individual peaks was confirmed by SDS-PAGE analysis (Figure 6.5B). The appearance of other proteins in the individual fraction lanes is the result of impure standards and the overlapping peaks caused by the resolution limits of the column as evident in the chromatograph. The protein peaks from the retention volume from the gel filtration were then plotted against the logarithm of the molecular weight of the protein standards (Figure 6.5C), from which a standards curve can be determined to allow molecular weight predictions from the retention volume of proteins.

On three separate gel filtration runs RPA70N, $PrimPol_{480-560}$, and the two proteins together were loaded onto the analytical gel filtration column and their retention volume measured (Figure 6.6A). The absorbance peak for $PrimPol_{480-560}$ is much greater than RPA70N, despite both proteins being loaded at the same volume and concentration (0.5 mL of 35 µM protein) and having a similar size because the extinction coefficient at 280 nm of RPA70N is 2980 mol^{-1} dm^3 cm^{-1} compared with $PrimPol_{480-560}$, which has an extinction coefficient of 8480 mol^{-1} dm^3 cm^{-1}. This difference is largely due to the presence of fewer tryptophan and tyrosine residues in RPA70N, which absorb UV light at a wavelength of 280 nm.

The $PrimPol_{480-560}$ has a much lower retention volume for a protein of its predicted size at 10.52 mL; this retention rate corresponds to a predicted molecular weight of ~42 kDa. The likely reason for this is that proteins are separated by their Stokes radius during size-exclusion chromatography. Our standards, however, have been calibrated for globular proteins by their molecular weight, instead of Stokes radius. It is probably unsurprising given that $PrimPol_{480-560}$ is predicted to be unstructured that it elutes at a higher apparent molecular weight as it likely has a greater Stokes radius than if it was well-ordered (Section 6.2).

There is a clear interaction between the two proteins, the peak from RPA70N elution peak at 12.8 mL has disappeared in the PrimPol-RPA70N interaction chromatograph and the main peak (10.15 mL) is distinct from the $PrimPol_{480-560}$ peak (10.52 mL). As

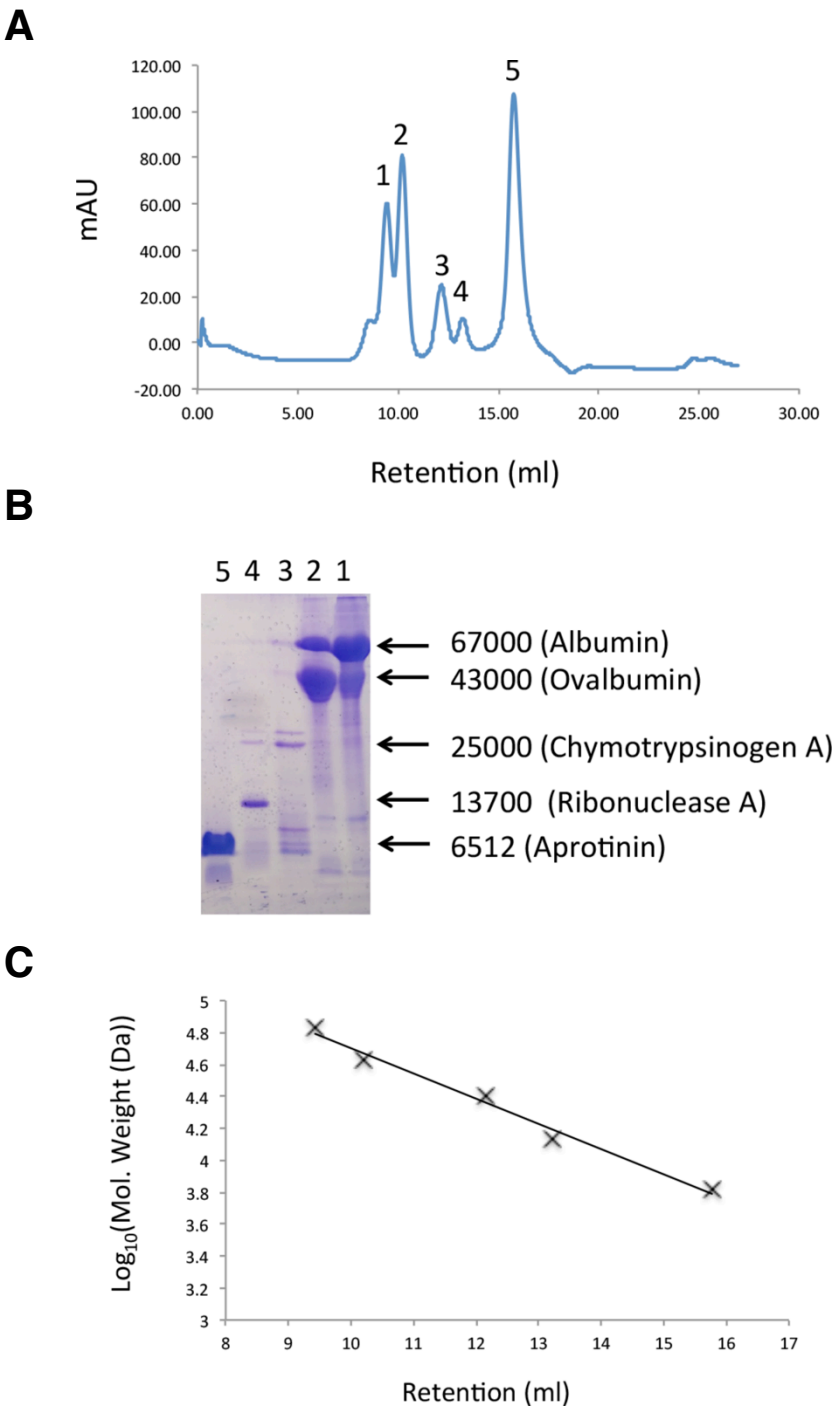


Figure 6.5. Calibration of analytical S75 gel filtration column.
A superdex S200 10/30 size exclusion chromatography column was calibrated through the use of protein standards. **(A)** Chromatograph showing the retention volumes of each of the proteins albumin, ovalbumin, chymotrypsinogen A, ribonuclease A and aprotinin. **(B)** Samples were taken at various points during the calibration denoted on the chromatograph as 1-5. These were analysed by SDS-PAGE and shown to correlate to the correct proteins as expected, corresponding molecular weight values are shown. Presence of more than one protein standard, particularly in fractions 1 and 2, are caused by overlapping peaks in the chromatograph due to low resolution. **(C)** The logarithmic molecular weight values are plotted against retention volume and a regression line plotted that can be used to calculate the predicted molecular weight of a complex given its retention volume.

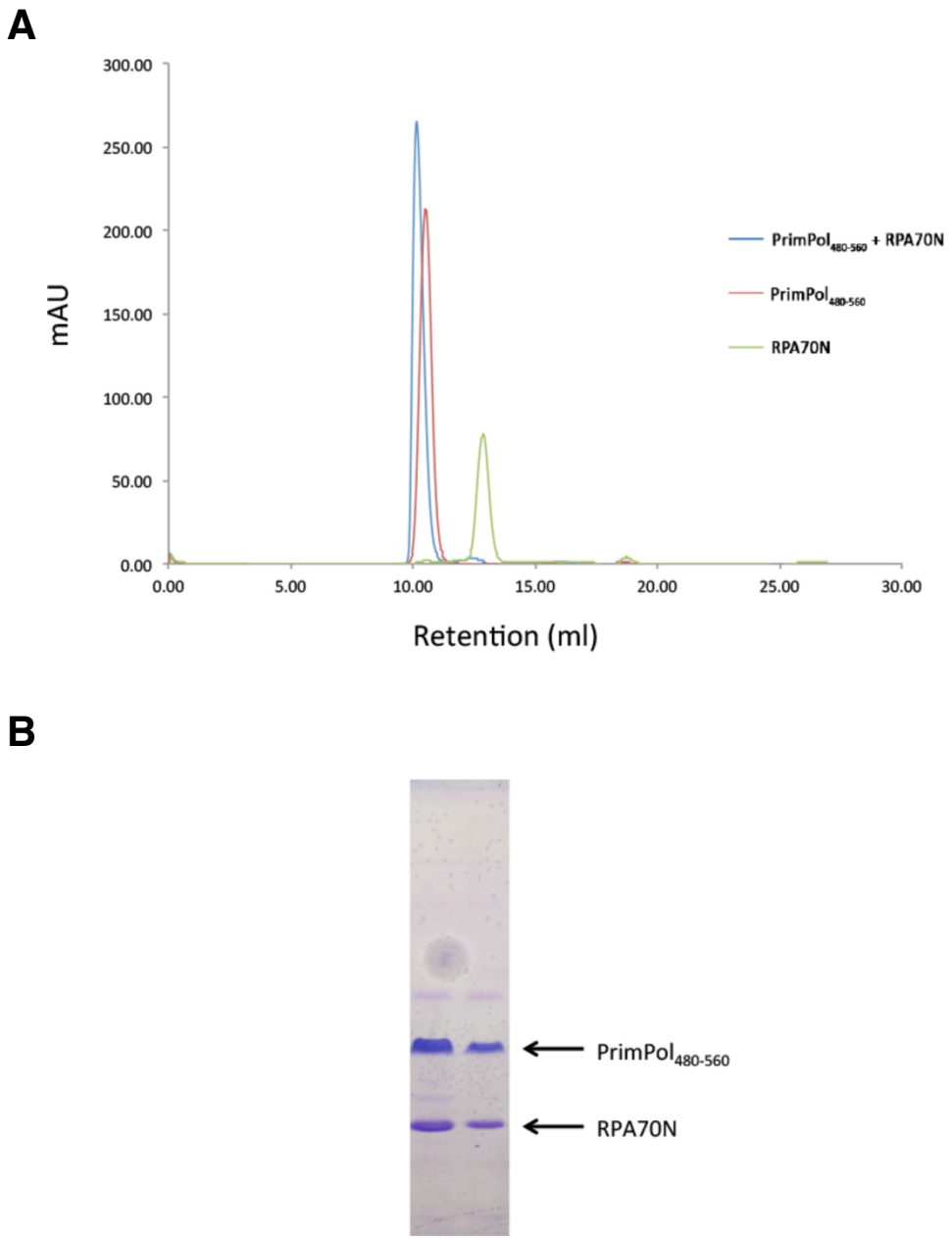


Figure 6.6. PrimPol$_{460-580}$ interacts with RPA70N.
(A) Chromatograph showing the retention volumes of PrimPol$_{480-560}$ (Red), RPA70N (Green) and PrimPol$_{480-560}$ and RPA70N in a 1:1 ratio (Blue). 0.5ml of 35µM protein was loaded in each case onto a superdex S75 10/30. The lower relative mAU of RPA to PrimPol$_{480-560}$ despite the same concentration and volume is the result of RPA70N having a lower molar extinction coefficient than PrimPol. RPA70N has a retention peak volume of 12.86 mL and PrimPol$_{480-560}$ has a retention peak volume of 10.52 mL. The PrimPol$_{460-580}$-RPA70N mixture chromatograph shows both individual proteins' peaks abrogated and a single peak at a retention volume of 10.15 mL, demonstrating an interaction between these two proteins. **(B)** Protein interaction was confirmed by SDS-PAGE analysis of the single peak from the PrimPol$^{460-580}$-RPA70N chromatograph, showing both proteins present on the gel in an approximately 1:1 ratio.

the two individual proteins peaks have disappeared completely and there is a single, shifted peak with an area approximately the same as the two peaks combined, there is likely to be an appreciable 1:1 interaction between the two proteins. This interaction was confirmed by SDS-PAGE of the fractions of the PrimPol-RPA70N peak (Figure 6.6B), which shows both $PrimPol_{480-560}$ and RPA 70N in the fraction.

The PrimPol-RPA70N peak corresponds to a predicted molecular weight of ~48 kDa. As the two proteins are found in the SDS-PAGE analysis to be present at an approximately 1:1 ratio, the protein complex could form a stable tetramer of $(RPA70N)_2(PrimPol_{480-560})_2$ with a predicted molecular weight of 48,398 Da. I, however, hypothesise that the greater Stokes radius of PrimPol480-560 caused by a lack of tertiary structure causes this peak. Regardless of whether the protein complex is a tetramer or a dimer, the evidence clearly indicates that the two proteins interact.

$PrimPol_{480-560}$ was isotopically labelled with ^{15}N and analysed for its HSQC spectrum by the Chazin group (Figure 6.7). The low dispersion observed in the HSQC spectrum is expected of a protein with little appreciative secondary structure. Upon titration of RPA70N at a 2:1 molar ratio, there is a remarkable effect on the HSQC spectrum, with peaks attenuating, broadening or shifting. Possible explanations for the causes of these effects include heterogeneous binding, motif swapping, and/or binding-induced partial folding of the domain in the complex.

6.4. Identifying RPA Interaction Motifs in Human PrimPol

Previous interaction studies with the RPA70N domain have identified a basic cleft into which several interacting partners bind, including RAD9, MRE11, ATRIP and p53 (Xu et al., 2008). These interacting partners all contain a similar sequence that they utilise to bind RPA70N (Xu et al., 2008), this sequence is highly negatively charged and is centred around an invariant pair of aspartate resiudes (Figure 6.8A). PrimPol also contains a potentially similar sequence towards its C-terminus (Figure 6.8A, amino acids 513-527) and our secondary structure prediction (Appendix I) predicts that, like with the other interacting partners, this portion of the protein may form an α helical structure. This motif of PrimPol is henceforth referred to as RPA-binding motif A (RBM-A).

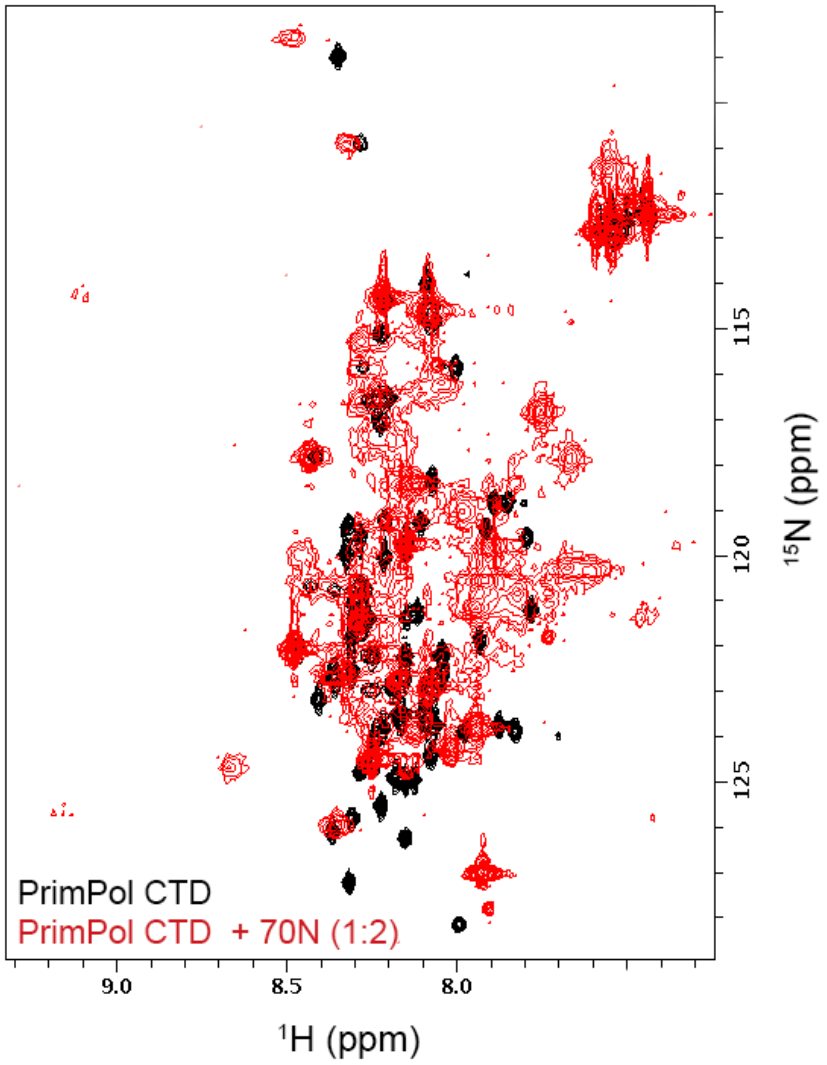
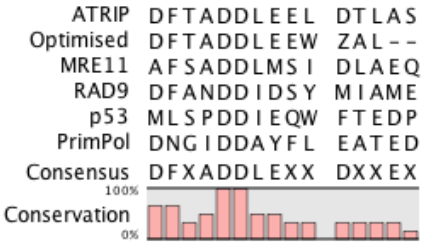


Figure 6.7. HSQC NMR confirmation of the PrimPol$_{460\text{-}580}$ interaction with RPA70N.
HSQC NMR spectra of ^{15}N-enriched PrimPol$_{460\text{-}580}$ (PrimPol CTD) in the absence (black) and presence (red) of RPA70N (70N) titrated in at a 2:1 ratio. Upon titration of RPA70N there is a significant alteration in the NMR spectrum, strongly suggestive of an interaction between PrimPol$_{460\text{-}580}$ and RPA70N. The experiments shown here were carried out by Dr. Aaron Ehlinger from the Chazin laboratory.

A



B

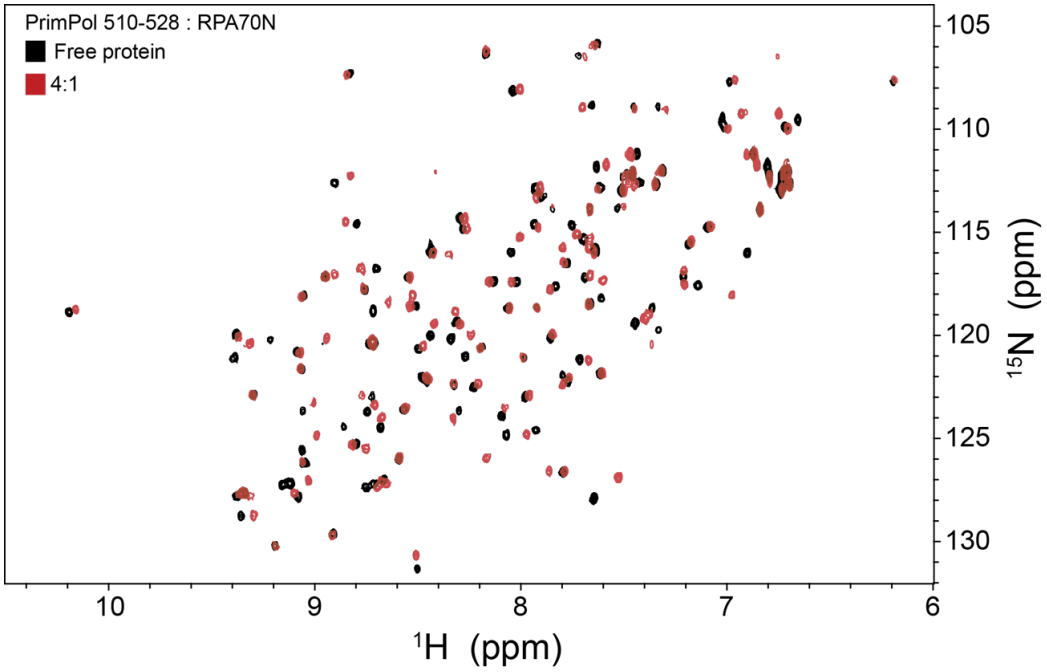


Figure 6.8. PrimPol contains a putative RPA70N-binding motif is shown to interact with RPA70N.

(A) The RPA70N binding motifs of ATRIP, MRE11, RAD9 and p53 have been identified, as well as an optimised RPA70N binding sequence used to obtain the structure of RPA in complex with a peptide helix, where Z is 3,4-Dichlorophenyl (Frank et al., 2014). PrimPol contains a sequence similar to this previously identified RPA70N-binding motif towards its C-terminus at amino acids 513-527. **(B)** $PrimPol_{510-528}$ interacts with RPA70N. HSQC NMR data was collected for ^{15}N-enriched RPA70N in the absence (black) or presence (red) of $PrimPol_{510-528}$, a short peptide sequence that contains the putative RPA70N-binding motif identified in (A). Locational shifts in the NMR spectra indicate that $PrimPol_{510-528}$ interacts with RPA70N.

To test whether RPA70N interacts with this putative interaction motif and how strong this interaction is, a 15 amino acid peptide containing RBM-A was synthesised (GenScript) and fluorescently labelled with the aim to carry out fluorescence anisotropy. However, upon testing anisotropy given an increase in RPA70N concentration, we found no changes in the apparent fluorescence polarisation (Appendix IV). Despite fluorescence labelling based on the manufacturers instructions and subsequent purification by size-exclusion chromatography (Superdex Peptide 10/300 GL column), ion exchange chromatography (mono-Q column), and reversed-phase chromatography (C18 column), the label that was present in the peptide samples was confirmed not to be covalently linked to the peptide by mass spectrometry.

Concurrently with our own attempted binding analysis, the Chazin group carried out HSQC NMR to acquire spectra of ^{15}N-enriched RPA70N in the absence (black) and presence (red) of PrimPol amino acids 510-528, containing RBM-A, at a 4:1 molar ratio to RPA70N (Figure 6.8B). There are significant amide chemical shifts evident in the location of the peaks and a weighted average of these shifts were plotted based on the difference between RPA70N alone and with the PrimPol peptide present (Figure 6.9A). The perturbations above a defined threshold ($\Delta\delta>0.1$) were mapped onto the structure of the RPA70N domain and compared with the perturbations caused by the binding of other RPA-interacting proteins ATRIP, Rad9 and MRE11 (Figure 6.9B)(Xu et al., 2008). This PrimPol peptide binds to a remarkably similar region to these other RPA70N protein partners, along the length of one side of the beta barrel structure and not with the helices flanking this domain. The perturbations map to the basic cleft of RPA70N (Figure 6.10), which is unsurprising given how the acidic nature of this peptide (theoretical pI=3.25).

6.5. Crystal Structure of the PrimPol-RPA70N Interaction

The X-ray diffraction data used to elucidate the structure of RPA70N revealed that the protein crystals conform to a hexagonal crystal lattice (Bochkareva et al., 2005). Unfortunately, the basic cleft to which RPA70N-interacting proteins bind is occluded from this crystal lattice and this prevents the binding to the interaction surface (Feldkamp et al., 2013). Three of the basic arginine residues from the cleft interacted with two glutamate residues (E7 and E100) from adjacent molecules. Three charge-reversal RPA70N variants were constructed by Feldkamp et al., E7R, E100R and an E7R/E100R variant that contained both mutations (Feldkamp et al., 2013). All three

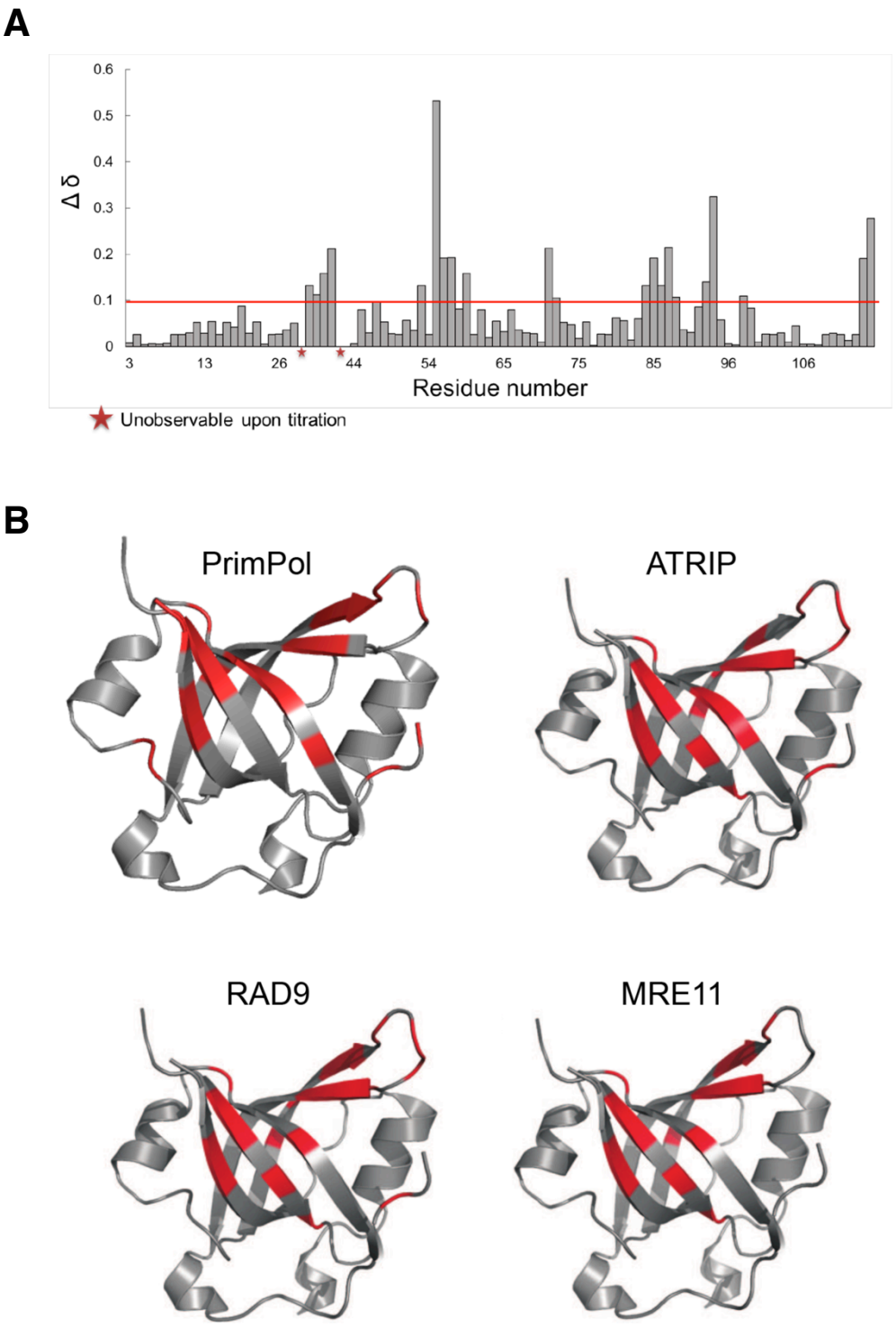


Figure 6.9. PrimPol interacts with RPA70N in the same region as other interacting proteins.
(A) Alterations in the HSQC NMR spectra following titration of $PrimPol_{510-528}$ were quantified and plotted based on the RPA70N residue number. **(B)** Perturbations above a Δδ threshold of 0.1 were mapped onto the structure of RPA70N (shown in red). The interacting region of PrimPol is comparable to that of other RPA70N-interacting proteins ATRIP, RAD9 and MRE11, binding mostly on one side of the beta-barrel.

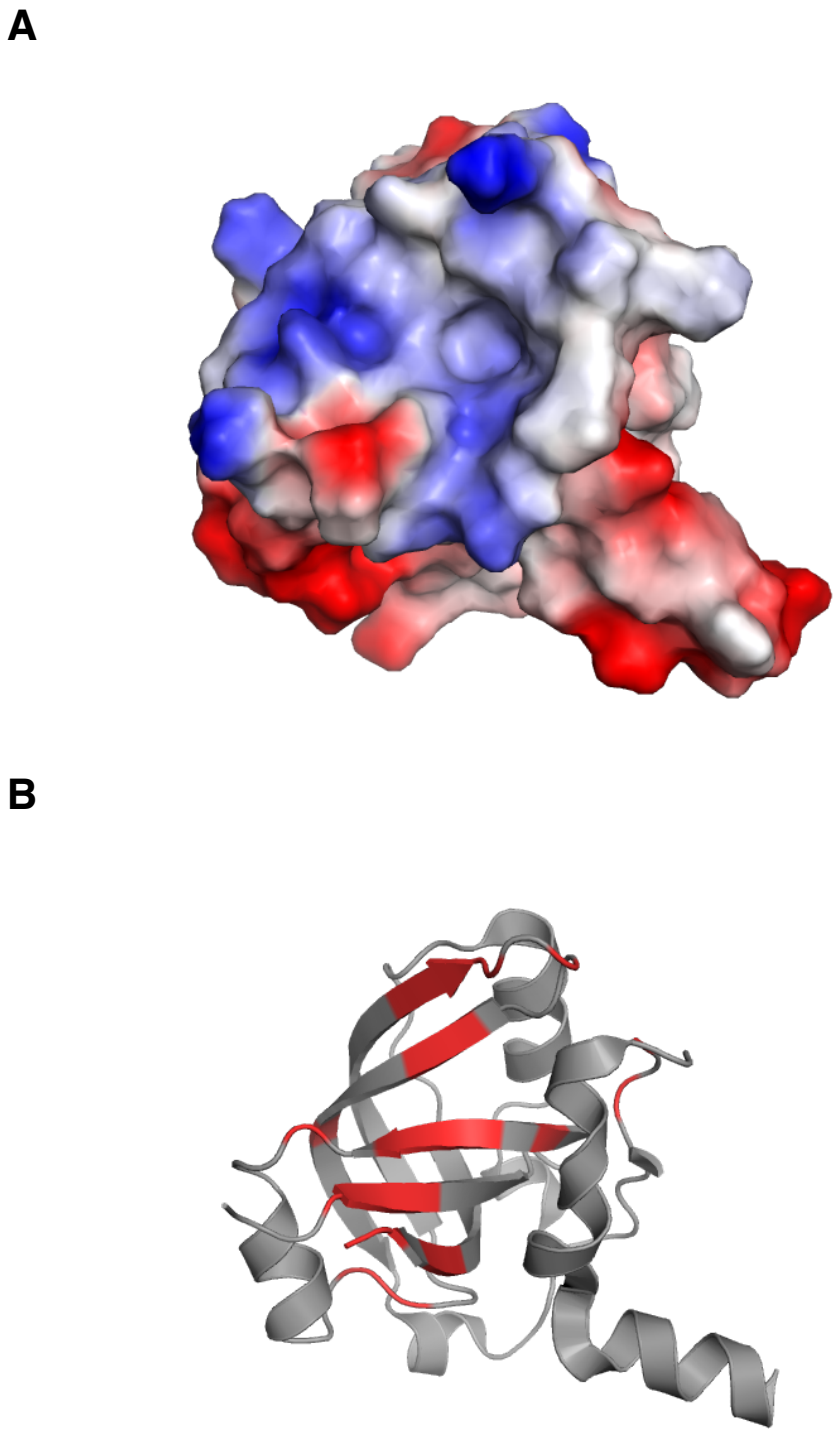


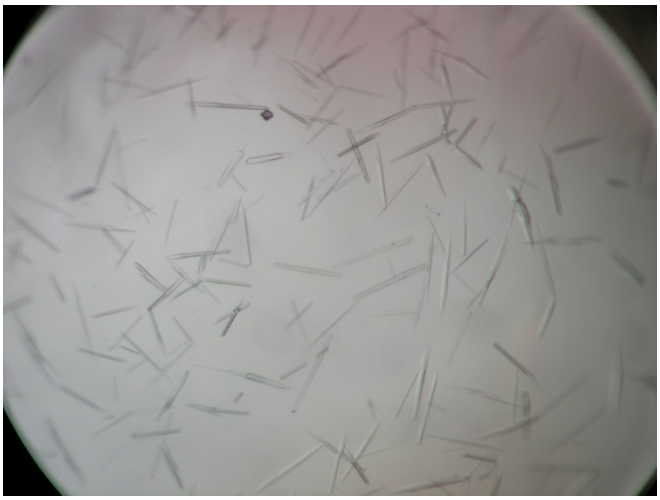
Figure 6.10. PrimPol binds at the basic cleft of RPA70N.
A surface model of RPA70N (top) is shown with the relative charge mapped onto the different regions of protein, with blue indicating positive charge and red indicating negative charge. The regions that PrimPol interact with are mapped onto the ribbon model of RPA70N (bottom) and a comparison of these regions illustrates that PrimPol binds to a basic cleft in RPA70N.

mutants successfully crystallised in a $P2_12_12_1$ orthorhombic lattice that made the basic cleft accessible. Their binding affinities for ATRIP were assessed to ensure that there is no significant changes in binding interaction partners in the basic cleft caused by these point mutants (Feldkamp et al., 2013). To demonstrate that structures of ligand complexes could be produced, a co-crystal structure of a binding cleft inhibitor with $RPA70N^{E7R}$ was obtained (Feldkamp et al., 2013). Subsequently, another $RPA70N^{E7R}$ crystal structure was obtained with an optimised high-affinity variant of the peptide helix of ATRIP that interacts with RPA70N in its basic cleft (Frank et al., 2014).

We aimed to obtain the structure of PrimPol peptides/domains bound to RPA70N. The RPA70N protein provided by the Chazin group was the E7R modified variant of the domain. To confirm that we were able to reproduce previous results, RPA70N was screened in the range of conditions that previously brought about crystals for any of E7R, E100R or both mutations and crystals formed after just 2 days. The crystals grown in 100mM Bis-Tris HCl (pH 5.5), 2M ammonium sulphate were chosen for collecting a full X-ray diffraction data set (ϕ angles 0°-359° taken in duplicate) (Figure 6.11). A molecular replacement using RPA70N as the search model matches the electron densities from our data. All of the structural studies described here were performed in joint collaboration with my colleague Dr. Nigel Brissett.

The 15 amino acid peptide containing RBM-A was chemically synthesised (sequence in Figure 6.8A), and incubated with $RPA70N^{E7R}$ and screened for co-crystallisation with 700µM $RPA70N^{E7R}$ at a 2.5:1 molar ratio. This concentration and peptide molar ratio was previously used to obtain co-crystals with the synthetic optimised ATRIP peptide (Frank et al., 2014). The proteins were screened using the Stura Footprint and MacroSol screens, which are optimised for the analysis of protein complexes (Stura et al., 1992). A number of conditions produced protein crystals suitable for X-ray diffraction analysis and four examples are shown in Figure 6.12.

The crystal lattice structures of $RPA70N^{E7R}$ alone and in complex with a helix in its basic cleft were significantly different, both in unit cell dimensions and lattice structures. Whilst $RPA70N^{E7R}$ crystals conform to a $P2_12_12_1$ orthorhombic lattice with 90° α, β and γ angles, in complex they crystallise in the C121 space group, forming a monoclinic crystal with 90° α and γ angles but a β angle of 120.6° (Feldkamp et al., 2013; Frank et al., 2014). $RPA70N^{E7R}$ unit cell a, b and c dimensions were 38.54, 53.72 and 54.36 Å



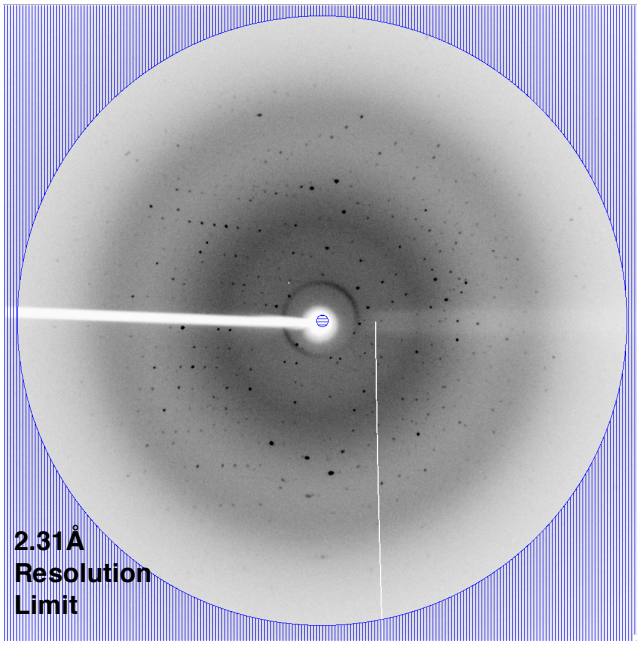


Figure 6.11. RPA70N^{E7R} readily forms protein crystals.
RPA70N^{E7R} forms crystals in a buffer of 100 mM Bis-Tris HCl (pH 5.5) and 2 M $AmSO_4$. A data set was taken in duplicate between ϕ angles 0°-359° and a representative image taken at a ϕ angle of 278° is shown with a 2.31Å resolution limit. A molecular replacement using RPA70N^{E7R} (PDB ID: 4IPC) as the search model matched our experimental data, confirming that these crystals were RPA70N^{E7R}.

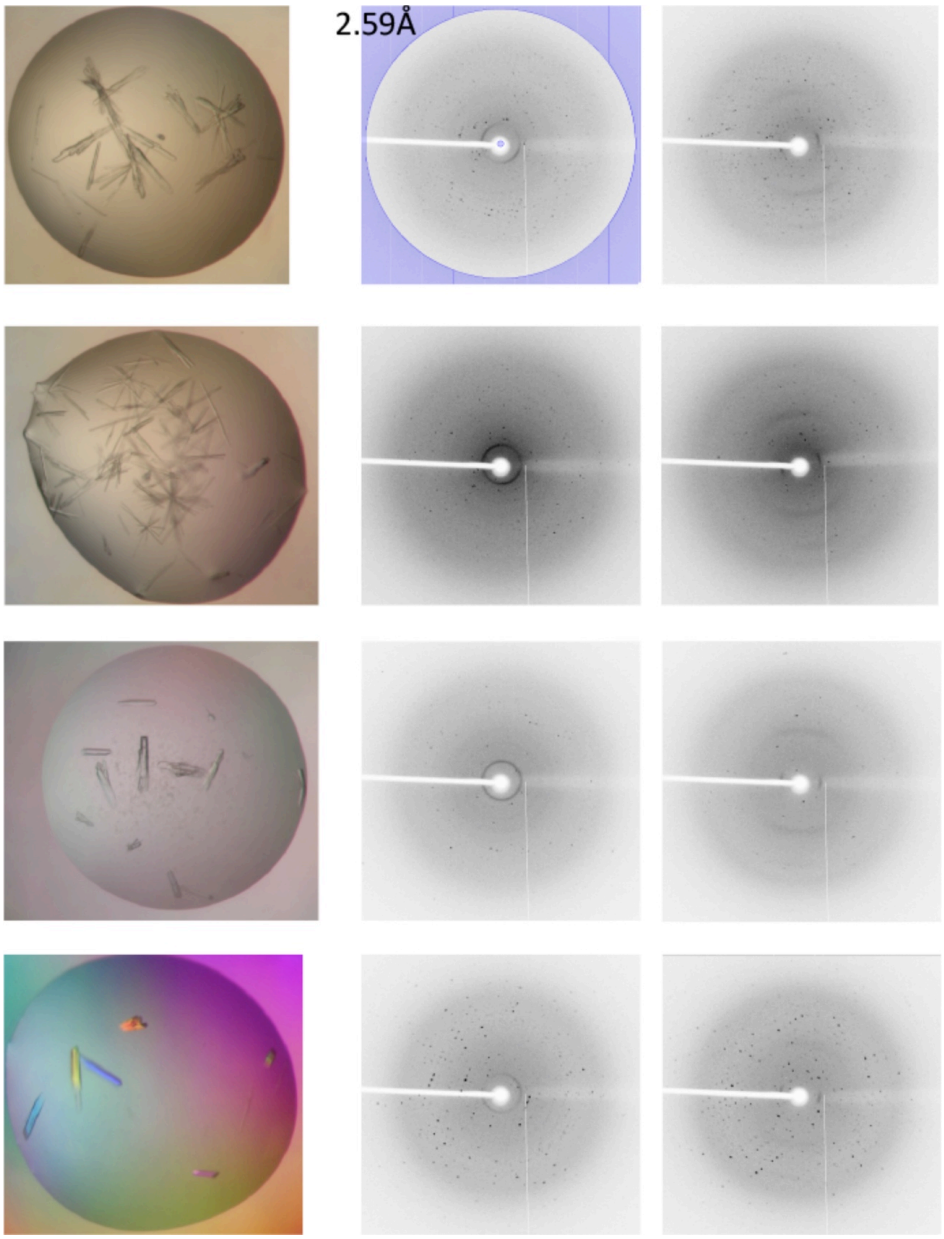


Figure 6.12. Co-crystallising RPA70N^{E7R} with PrimPol$_{513-527}$.
RPA70N^{E7R} was screened for cocrystals with a 15 amino acid PrimPol peptide (PrimPol$_{513-527}$) 1:2.5 molar ratio (700 μM with 1750 μM peptide). Representative crystals and diffraction patterns taken at ϕ angles of 0° and 90° are shown. The four crystals shown were grown in buffers of 1.6 M NaH^2PO^4/K^2HPO^4 (pH7); 2 M NaH_2PO_4/K^2HPO^4 (pH7); 2% w/v PEG 8000 , 1.0 M lithium sulphate; and 12% w/v PEG 4000, 5% v/v 2-propanol , 0.1 M sodium citrate (pH5.5). A representative 2.59Å resolution limit is shown on the first diffraction pattern image.

respectively, whilst the protein complex crystals had unit cell dimensions of 90.15, 46.80 and 78.79 Å, respectively. As such, when indexing our crystals, we would expect to see a significant shift in the crystal lattice structure if the PrimPol peptide is present in the basic cleft. However, we found that there were no significant changes in the space group or unit cell dimensions of the crystals that were obtained (Table 6.1).

Alongside these studies with the 15 amino acid peptide of PrimPol (resiudes 513-528), we also screened for co-crystals of RPA70N^{E7R} bound to PrimPol$_{480-560}$. PrimPol$_{480-560}$ and RPA70N^{E7R} were incubated in a 1:1 ratio and then the mixture was gel filtered to allow the bound complex to be purified (see Section 6.3). This complex was set up in the Stura Footprint and MacroSol crysallisation screens, following a pre-crystallisation test (PCT) to determine a suitable concentration (9 mg mL^{-1}). A range of conditions produced crystals, seven of which are shown in Figure 6.13. Crystals from each of these conditions were indexed using iMOSFLM and again there were no significant changes in the space group or unit cell dimensions (Table 6.2).

Although there were significant crystal lattice structural changes in the presence of the α-helical peptide, RPA70N^{E7R} was crystallised in complex with an inhibitory compound (VU079104) in its basic cleft with a $P2_12_12_1$ orthorhombic lattice unit cell dimensions of 38.27, 54.27 and 54.33 Å. Therefore, despite no significant changes in the space group or unit cell dimensions, we collected data sets for each to determine if any of these crystals exhibited any additional electron density in the binding cleft. A molecular replacement, using RPA70N^{E7R} (PDB ID: 4IPC) as the search model, did not reveal any unaccounted electron density in the basic binding cleft for the first six conditions. However, in the condition containing 0.2 M imidazole malate (pH 6.0) and 30 % (w/v) PEG 4000 there was apparent additional electron density in this binding cleft that, although weak, was connected and therefore unlikely to be salt or water molecules. These crystals were sent for additional analysis using synchrotron radiation at the Diamond Light Source, Oxfordshire; synchrotron radiation produces more intense and monochromatic high-energy X-ray beams to give greater amplitude of signal and greater penetrance. The resolution obtained from the data at the synchrotron was ~1.3 Å (See diffraction data in Figure 6.14 and data collection and refinement statistics in Table 6.3).

Condition	Spacegroup	a	b	c	α	β	γ
0.15 M sodium citrate (pH5.5), 2 M ammonium sulphate	$P2_12_12_1$	38.0	53.3	53.8	90	90	90
2.0 M NaH_2PO_4/K_2HPO_4 (pH7.0)	$P2_12_12_1$	38.4	53.3	53.7	90	90	90
1.6 M NaH_2PO_4/K_2HPO_4 (pH7.0)	$P2_12_12_1$	38.8	53.2	53.8	90	90	90
0.01 M sodium borate (pH 8.5), 1.5 M sodium citrate	$P2_12_12_1$	39.8	52.5	53.6	90	90	90
0.01 M sodium cacodylate (pH 6.5), 45% (w/v) PEG 2000 MME	$P2_12_12_1$	39.1	50.8	52.2	90	90	90
1.0 M ammonium dihydrogen phosphate, 0.1 M sodium citrate (pH 5.5)	$P2_12_12_1$	38.2	53.1	54.5	90	90	90
2% w/v PEG 8000, 1.0 M lithium sulphate	$P2_12_12_1$	38.2	53.0	53.6	90	90	90
2% w/v PEG 8000, 1.0 M lithium sulphate, 0.1 M imidazole malate (pH 6.5)	$P2_12_12_1$	38.9	53.3	53.9	90	90	90
12% w/v PEG 4000, 5% v/v 2-propanol, 0.1 M sodium citrate (pH5.5)	$P2_12_12_1$	38.2	53.2	53.7	90	90	90
RPA70N E7R	$P2_12_12_1$	38.5	53.7	54.4	90	90	90
RPA70N E7R complex with optimised ATRIP helix	C121	90.2	46.8	78.8	90	121	90

Table 6.1. Crystal lattice properties of crystals of RPA70N^{E7R} with PrimPol$_{513-527}$.
All crystals exhibit a $P2_12_12_1$ orthorhombic lattice and similar lattice dimensions suggesting that the small peptide is not present in the RPA70N^{E7R} crystals.

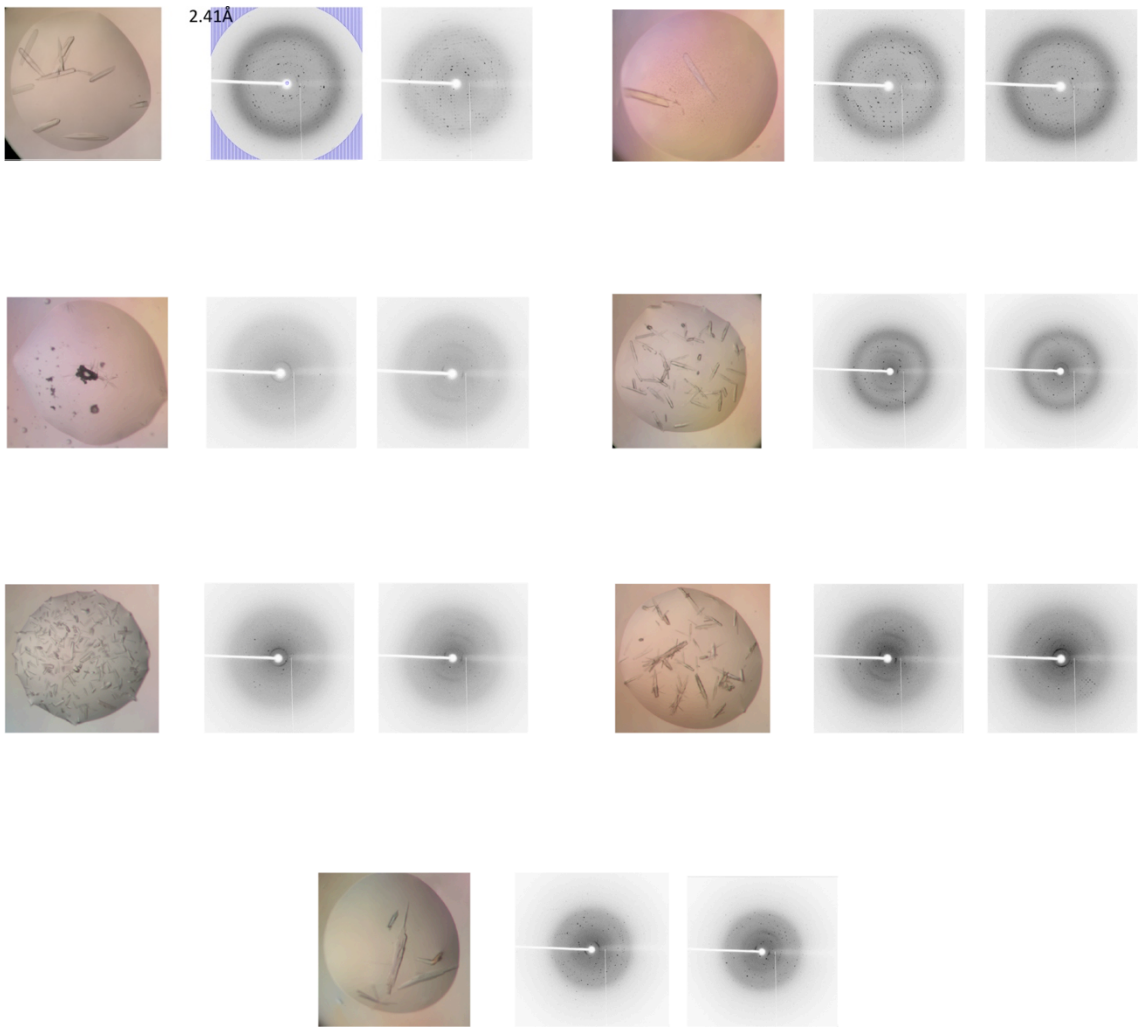
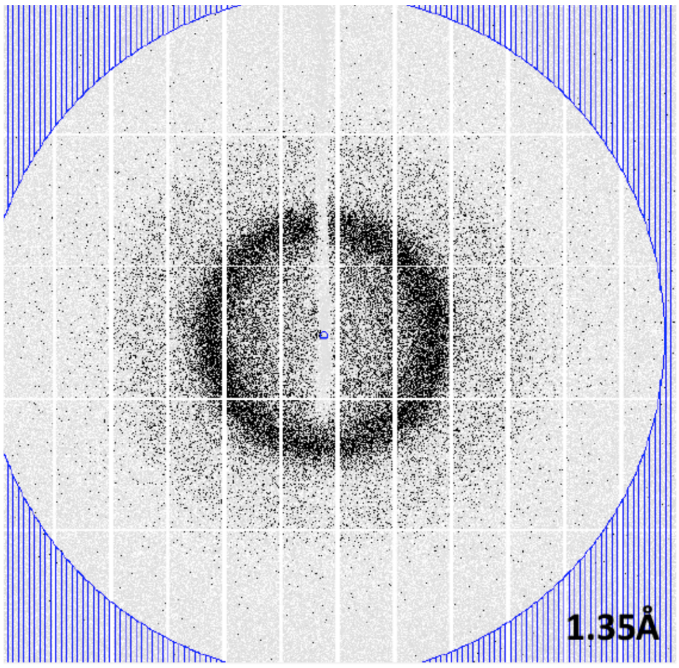


Figure 6.13. Co-crystallising RPA70N^{E7R} with PrimPol$_{480-560}$.
RPA70N^{E7R} and PrimPol$_{480-560}$ mixed at 1:1 ratio, then gel filtered on an analytical gel filtration column; the fractions from the main peak in gel filtration were then screened for crystals at 9 mg ml^{-1}. Full Data sets were collected and indexed for each of these conditions and representative image taken at a ϕ angles of 0° and 90°. Crystals were grown in buffers of **(A)** 0.2M Imidazole malate pH 5.5, 33% (v/v) PEG 600; **(B)** 0.1M Na Hepes (pH 8.2), 40% (v/v) PEG 550 MME; **(C)** 0.2 M ammonium acetate, 0.1 M sodium acetate (pH 4.5), 20% w/v PEG 3350; **(D)** 0.15 M Sodium citrate (pH 5.5), 1.5M Ammonium Sulphate; **(E)** 1.32 M NaH_2PO_4/K_2HPO_4 pH 7.0; **(F)** 0.01M Sodium Borate pH 8.5, 1.2M Sodium Citrate and **(G)** 0. 2 M imidazole malate (pH 6.0), 30 % (w/v) PEG 4000.

Condition	Spacegroup	a	b	c	α	β	γ
0.2 M Imidazole malate (pH 5.5), 33% (v/v) PEG 600	$P2_12_12_1$	39.2	51.0	52.5	90	90	90
0.1 M Na Hepes (pH 8.2), 40% (v/v) PEG 550 MME	$P2_12_12_1$	38.2	52.8	53.7	90	90	90
0.2 M ammonium acetate, 0.1 M Na acetate (pH 4.5), 20% (w/v) PEG 3350	$P2_12_12_1$	37.8	52.9	54.5	90	90	90
0.15 M Sodium citrate (pH 5.5), 1 M Ammonium Sulphate	$P2_12_12_1$	38.4	53.5	54.1	90	90	90
1.32 M NaH_2PO_4/K_2HPO_4 (pH 7.0)	$P2_12_12_1$	38.9	52.5	52.0	90	90	90
0.01 M Sodium Borate (pH 8.5), 1.2 M Sodium Citrate	$P2_12_12_1$	38.3	53.2	53.9	90	90	90
0. 2 M imidazole malate (pH 6.0), 30 % (w/v) PEG 4000	$P2_12_12_1$	38.0	53.5	53.9	90	90	90

Table 6.2. Crystal lattice properties of crystals of RPA70N^{E7R} with PrimPol$_{480-560}$.
All crystals exhibit a $P2_12_12_1$ orthorhombic lattice and similar lattice dimensions. However, following indexing and molecular replacement, unaccounted electron density was evident in the putative binding pocket of RPA70N for crystals grown in 0.2 M imidazole malate (pH 6.0), 30 % (w/v) PEG 4000.



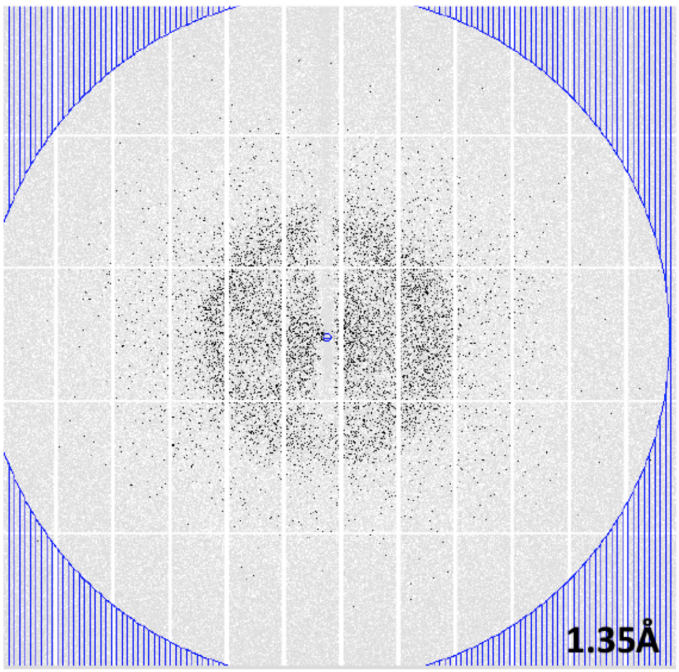


Figure 6.14. Diffraction pattern for RPA70N^{E7R} in complex with PrimPol$_{480-560}$. Diffraction pattern is shown from the synchrotron at the Diamond Light Source in Oxfordshire for the crystals of RPA70N^{E7R} in complex with PrimPol$_{480-560}$ grown in a buffer of 0.2 M imidazole malate (pH 6.0), 30 % (w/v) PEG 4000. Full Data sets were collected and indexed and representative images shown are at a ϕ angles of 0° and 45°. A 1.35Å diffraction limit is shown.

Data Collection	
Source	Station I03, Diamond
Space group	$P2_12_12_1$
Unit cell dimensions	
a, b, c (Å)	38.05, 53.49, 53.90
α, β, γ (°)	90.00, 90.00, 90.00
Wavelength (Å)	0.91406
Resolution Range (Å)	16.25 (1.28)
Total number of observations	352183
Number of unique observations	29019
Overall I/(σI)	26.9 (2.7)
Overall completeness (%)	99.8 (98.2)
R-merge	0.044 (0.655)
Redundancy	12.1 (7.4)
Refinement	
Resolution (Å)	16.25 (1.28)
Number of reflections	27557
R_{factor}/R_{free}	0.238/0.192
Contents of assymetric unit	2 peptides
Number of non-hydrogen atoms	1070
Protein	1021
Water molecules	49
Mean B value ($Å^2$)	17.467
RMSDs	
Bonds (Å)	0.021
Angles (°)	2.072
Ramachandran Statistics	
Favoured regions (%)	97.32
Allowed regions (%)	0.89
Disallowed regions (%)	1.79

Table 6.3. Data collection and refinement statistics for RPA70N^{E7R} in complex with PrimPol$_{480-560}$.
Values in parentheses refer to those in the highest resolution shell. RMSD = Root-mean-square deviation.

Following model building into this density and refinement of the crystal structure by Dr. Nigel Brissett it was evident that there was a helical peptide that was present in the basic binding cleft of RPA70N (Figure 6.15). Unexpectedly, it was not the peptide sequence that was previously identified that was found to be interacting with the RPA70N binding site, RBM-A. Instead, the sequence that is bound is in fact VDEIPDELIIEVLQE at the very C-terminus of PrimPol (amino acids 545-560). This C-terminal motif is henceforth referred to as protein binding motif B (RBM-B). As the amino acids towards the N- and C-terminus of this motif were low resolution as the flexibility increases towards the end of the α-helical structure, only amino acids 547-557 are shown (EIPDELIIEV).

RBM-B, like RBM-A (see Section 6.4), binds between the β-barrel into the basic cleft in RPA70N (Figure 6.15). RBM-B also has a low pI, predicted to be 3.34, so it is perhaps unsurprising that it binds into the basic cleft of RPA70N. There is no significant change in the structure of RPA70N upon binding of RBM-B. Despite the amino acids 480-544 not being visible in our area of contiguous electron density, the helix that is bound, RBM-B, shows a significantly high goodness of fit to the electron density that is evident (Figure 6.16A) for amino acids 548-557. As such, this data is reliable in showing the binding region of PrimPol. We expect that residues 480-544, because they are not bound to RPA are flexible and this temporal protein disorder does not allow for resolution by X-ray crystallography.

The helical orientation is stabilised by a number of electrostatic interactions between RBM-B and RPA70N (Figure 6.16B). The aspartate at position 551 of PrimPol is perhaps the most important point of contact as it interacts with the two arginines of RPA (R31 and R43) and a threonine (T34) side chain, as well as the amine group of the protein backbone of T34. The carbonyl group of PrimPol's isoleucine at position 549 likely acts as a hydrogen bond acceptor for the RPA's R43. The glutamate at position 548 forms an electrostatic interaction with an arginine (R93) on the other side of RPA70N's β-barrel, acting to secure the helix of PrimPol in this orientation. These electrostatic interactions are of paramount importance in the binding of PrimPol to RPA70N. In addition to the electrostatic interactions shown in Figure 6.16B, there are a number of non-bonding contacts and these are detailed in the Lig plot shown in Figure 6.17.

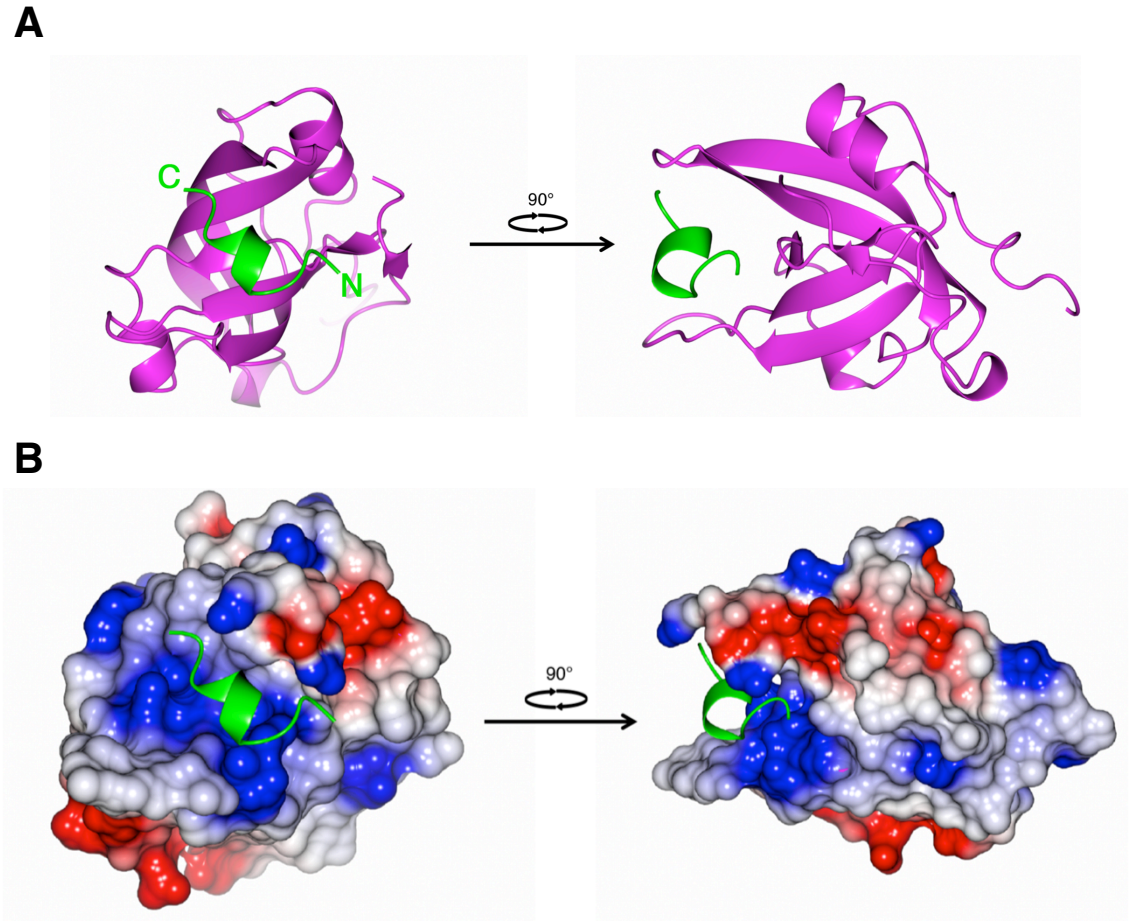


Figure 6.15. Co-crystal structure of RPA70N^{E7R} with a PrimPol α helix shows interaction of RBM-B with the basic cleft of RPA70N. **(A)** PrimPol PBM-B (green) binds to a cleft between β sheets in the β barrel of RPA70N (purple) **(B)** The acidic PrimPol PBM-B peptide binds to a basic cleft in the OB fold of RPA70N. The electrostatic surface model of RPA70N is shown in which blue indicates a basic surface and red indicates an acidic surface.

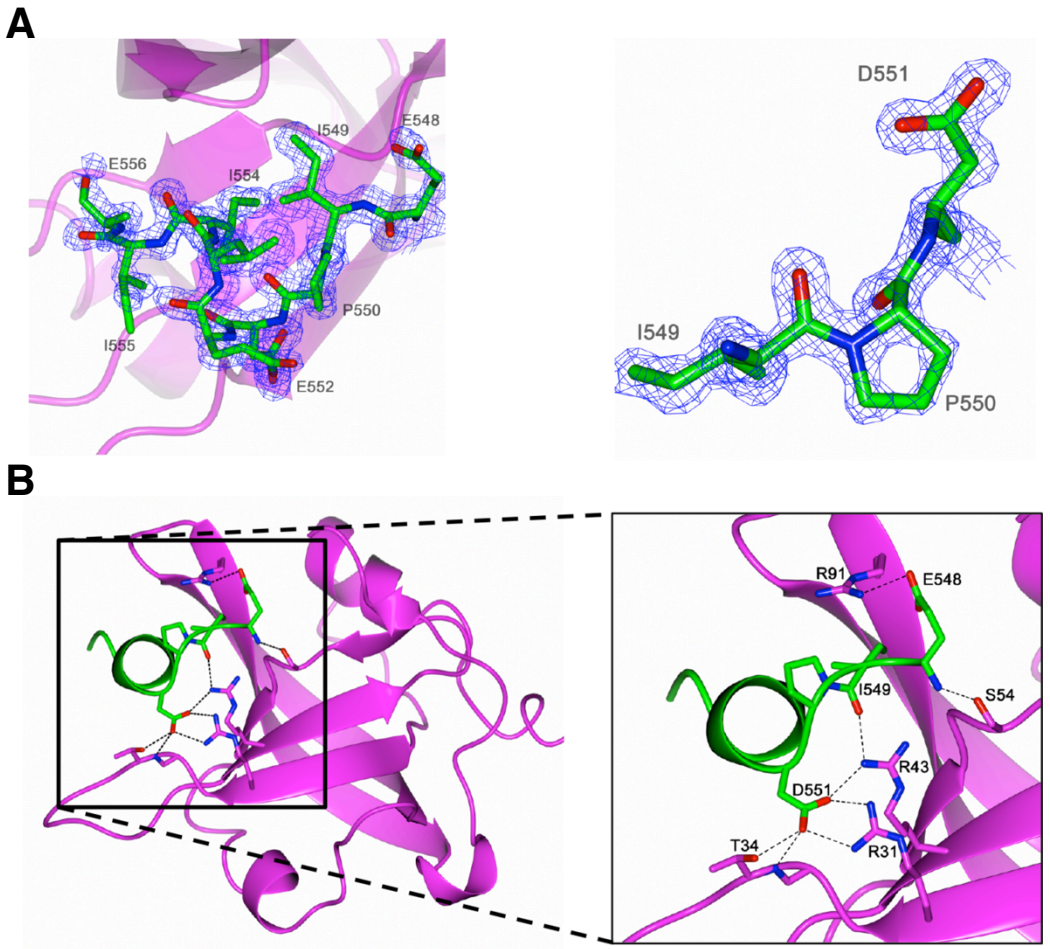


Figure 6.16. Electron density and protein-protein interaction contacts of RBM-B with RPA70N.
(A) The continuous electron density of the RBM-B PrimPol peptide is shown in its binding region of RPA70N. The electron density of three continuous residues (I549-D551) is also shown and clearly illustrates the resolution and quality of data, with the electron density clearly showing the ring structure of P550. **(B)** The electrostatic interactions of PrimPol with RPA70N. D551 is of particular importance as it forms a number of electrostatic interactions with both the side chains and an amine of the backbone of the RPA70N peptide.

Strikingly, the interaction between PrimPol RBM-B and RPA70N is significantly different from the interaction between RPA70N and either the modified ATRIP stapled peptide and the p53 peptide that have been described previously (Bochkareva et al., 2005; Frank et al., 2014). A superposition of the modified ATRIP peptide with RBM-B shows that the two helices bind in a similar region to RPA70N but they are in opposite orientations (Figure 6.18A). The main interaction of the modified ATRIP peptide is of its modified 3,4-dichlorophenyl amino acid into a hydrophobic pocket on RPA70N, and in p53 there is a phenylalanine residue that extends into this pocket. This pocket is also the region that the VUO79104 RPA70N binding inhibitor was bound in its co-crystal structure (Feldkamp et al., 2013). PrimPol RBM-B has an isoleucine (I554) that occupies the hydrophobic pocket on RPA70N (Figure 6.18B). I554 forms hydrophobic non-bonding contacts with a methionine (M57) and a valine (V93) of RPA70N in this pocket. However, these hydrophobic interactions are not the major contacts in the RPA70N-PrimPol RBM-B interface.

Despite having a pair of conserved aspartates, the stapled helix of ATRIP and the interacting helix of p53 do not have any equivalent electrostatic interactions to that of PrimPol RBM-B's D551. It required a number of unsuccessful attempts to crystallise the peptide of p53, the helix (p53 amino acids 33-60) was added as a covalent modification to RPA70N as a fusion peptide at the C-terminus of RPA70N. Also there was a number of modifications to the ATRIP peptide that did not successfully crystallise with RPA70N prior to the high affinity 3,4-dichlorophenyl interaction. The additional stabilising electrostatic interactions between PrimPol RBM-B and RPA70N provided by D551 could be the reason we were able to get a structure of RBM-B bound to RPA70N.

6.6. PrimPol has Two Protein Interaction Motifs

As the HSQC work from the Chazin laboratory had identified RBM-A as a RPA70N-interaction motif and our crystallography work had identified RBM-B as a RPA70N-interaction motif, we decided to check which of the two, if either, interacted with RPA70N in gel filtration experiments if the other was knocked out. To do this, two PrimPol variants were constructed by Dr. Nigel Brissett. The first, $PrimPol_{RMB\text{-}A\text{-}KO}$, is the C-terminal 80 amino acids of PrimPol (480-560), but with three charge reversal mutations and a single charge abrogation mutation (D514R, D518R, D519R and D528A), with the intended effect of removing the interaction between RBM-A and

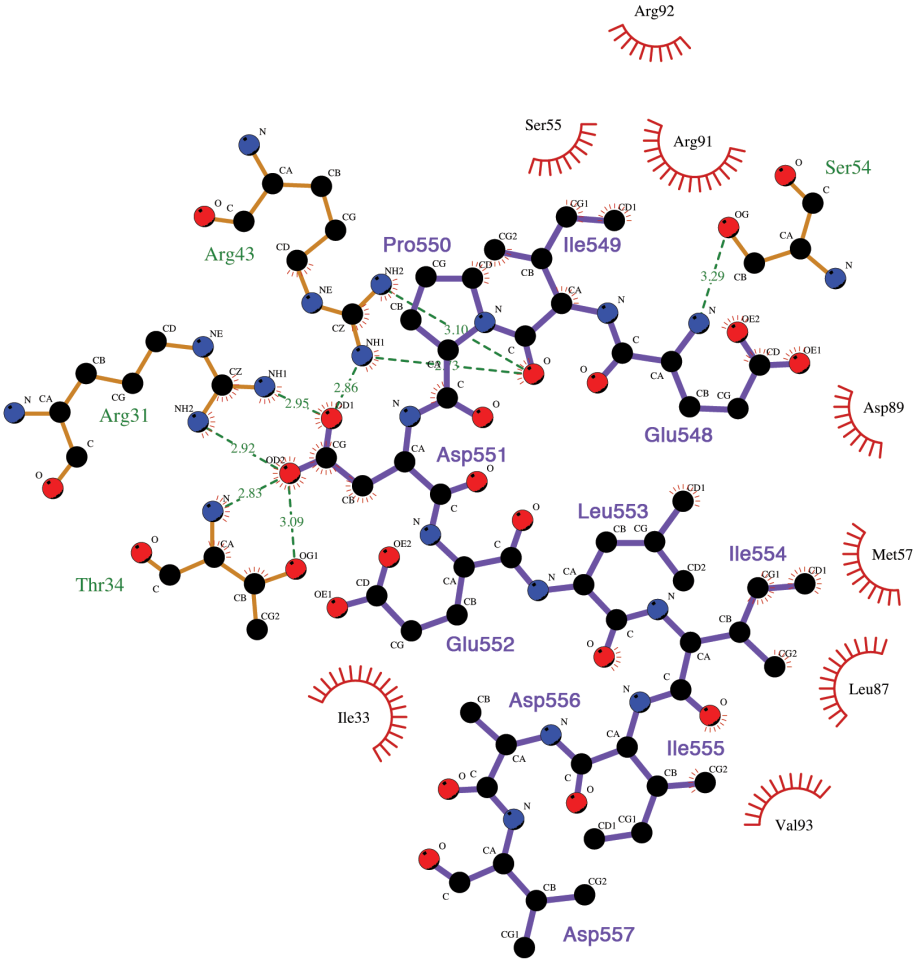


Figure 6.17. LigPlot showing the molecular interactions between RBM-B and RPA70N.

The electrostatic and non-bonding contacts between PrimPol RBM-B (purple residues) and RPA70N (brown residues) are shown in this LigPlot schematic, calculated using the LigPlot+ software (Lasowaski and Swindells, 2011).

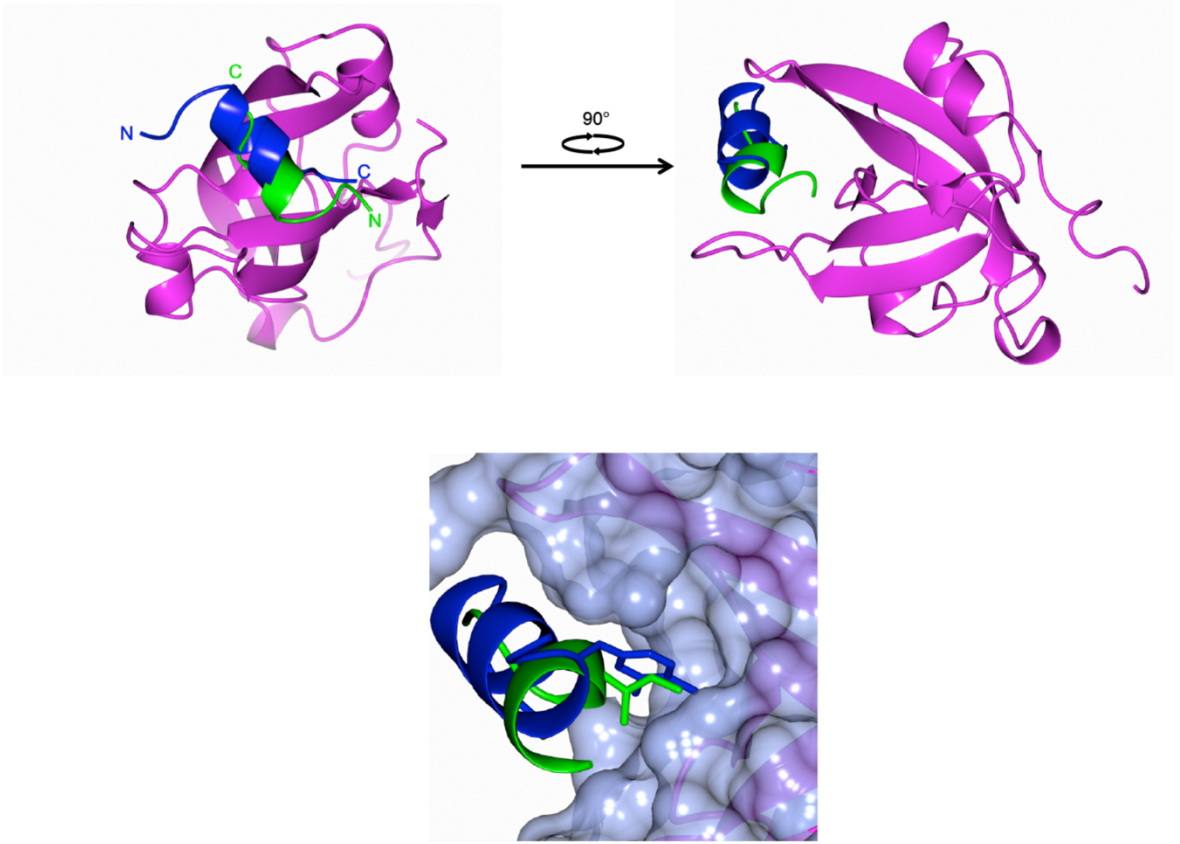


Figure 6.18. Comparison of the PrimPol RBM-B helix bound to RPA70N with a stapled ATRIP peptide.
(A) The PrimPol RBM-B helix (green) binds the same region of RPA70N (purple) as the stapled ATRIP peptide (blue; PDB ID: 4NB3), however the two peptides are bound in opposite orientations. **(B)** Binding to a hydrophobic pocket in RPA70N. The major interaction of the stapled peptide previously shown to bind RPA70N is of the modified 3,4-dichlorophenyl residue to a hydrophobic pocket in RPA70N. PrimPol PBM-B has an isoleucine (I554) that extends into this pocket but the major interactions are the electrostatic interactions shown in Figure 6.#. Amino acids 548-550 of PrimPol are not shown in (B) for clarity.

RPA70N. The second is PrimPol$_{480-546}$ with the C-terminal 14 amino acids removed from PrimPol to remove the RBM-B interaction with RPA70N. These proteins were purified as described in Table 2.5 (Figure 6.19).

These variants were analysed for their interaction with RPA70N by gel filtration in the same manner as PrimPol$_{480-560}$ (Section 6.3.4). Like PrimPol$_{480-560}$, both PrimPol$_{RBM-A-KO}$ and PrimPol$_{480-546}$, elute as a single defined multimeric complex from the chromatography column (Figure 6.20). When added in a 1:1 molar ratio with RPA70N, PrimPol$_{RBM-A-KO}$ and PrimPol$_{480-546}$ both interact with RPA70N. Whereas the RPA70N peak is completely removed upon binding to PrimPol for PrimPol$_{480-560}$, both PrimPol$_{RBM-A-KO}$ and PrimPol$_{480-546}$ show evidence of some unbound RPA70N with a small peak at the same elution volume as RPA70N. These small peaks were shown to indeed be free, unbound RPA70N by SDS-PAGE analysis (result not shown). This suggests that whilst both PrimPol$_{RMB-A-KO}$ and PrimPol$_{480-546}$ retain RPA binding activity, this binding is not as strong as PrimPol$_{480-560}$, which contains both RPA-binding motifs. A more quantitative analysis of the binding activity is required to determine which, if either, motif has higher affinity for binding RPA70N.

To ensure that these are the only RPA-binding motifs in PrimPol$_{480-560}$, two controls were constructed to assess whether a double RBM-A and RBM-B knockout would bind RPA. The first is a combination of the RBM-A and RBM-B knockouts above, PrimPol$_{480-546/RBM-A-KO}$. The second is a 6 point mutation variant (PrimPol$_{RBM-A-KO/RBM-B-KO}$; D514R, D518R, D519R, D551R, I554A, I555A). Both of these double RBM-A/RBM-B knockout variants of PrimPol have their RPA-binding capacity completely abrogated (Figure 6.21). There is no shift in the peaks of the individual proteins' size exclusion chromatograph peaks and the RPA peak remains intact.

6.7. RBM-A and RBM-B Represent Common Protein Interaction Motifs

Both RBM-A and RBM-B are conserved across PrimPol orthologues (Figure 6.22). Whilst RBM-B remains conserved down to inverterbrate species such as the sea urchin *Strongylocentrotus purpuratus* as shown, RBM-A splits into two parts in which the highly conserved double aspartic acid and the conserved C-terminus of RBM-A move away from each other in the primary sequence (for example, compare human DNGIDDAYFLEATED with *S. purpuratus* GYKSDDHLVPPEHPSFTDWDE). However,

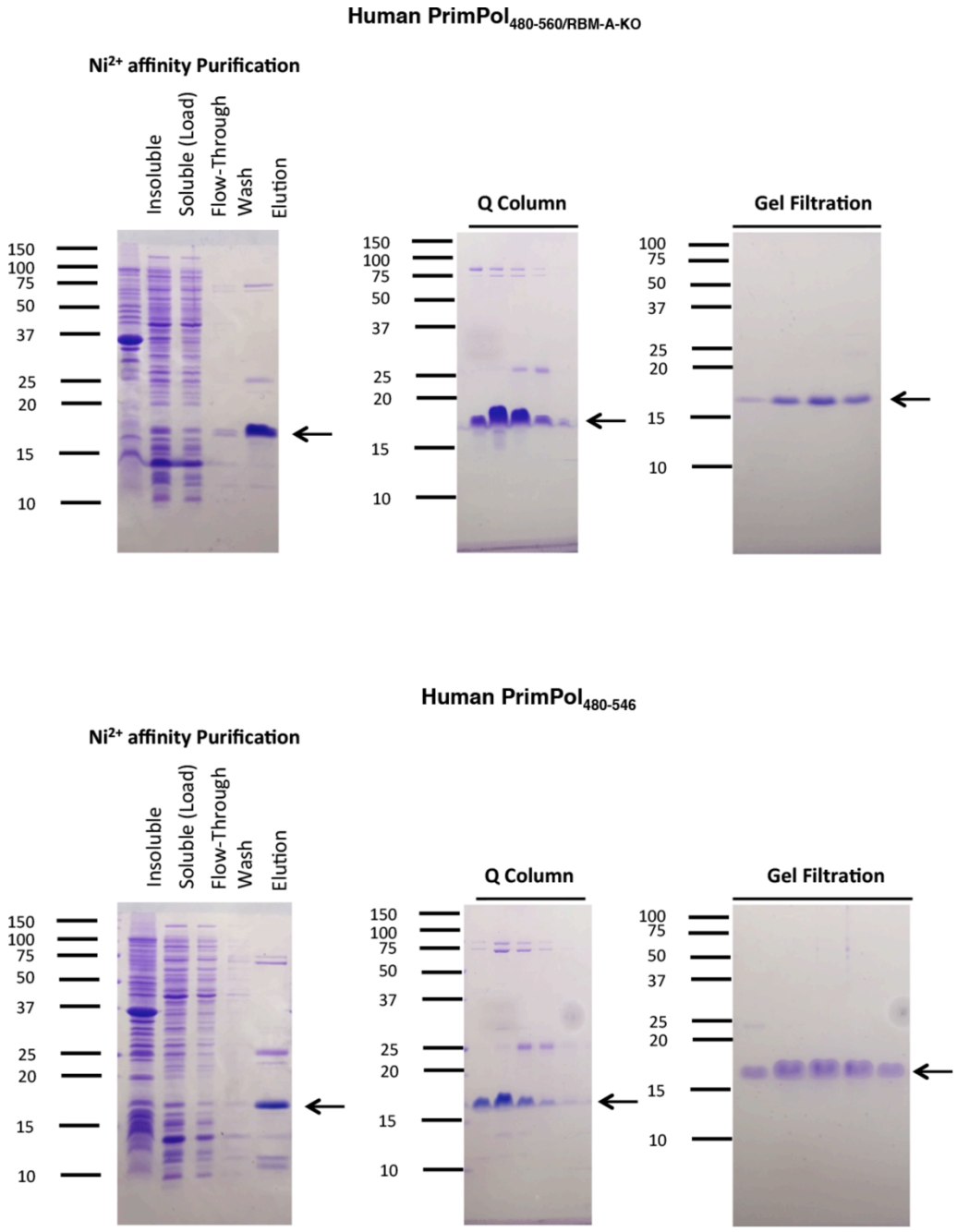


Figure 6.19. Purification of PrimPol$_{480\text{-}560/RBM\text{-}A\text{-}KO}$ and PrimPol$_{480\text{-}560}$.
(A) Purification of PrimPol$_{480\text{-}560/RBM\text{-}A\text{-}KO}$ (D514R, D518R, D519R, D528A; 11.1 kDa) by IMAC (Ni^{2+}) affinity, anion exchange and size exclusion chromatography. **(B)** Purification of PrimPol$_{480\text{-}546}$ (9.4 kDa) missing RBM-B by IMAC (Ni^{2+}) affinity, anion exchange and size exclusion chromatography. SDS-PAGE analysis of samples from purification fractions of both proteins indicate a protein of a size between 15 and 20 kDa. As this is lower than PrimPol$_{480\text{-}560}$ (runs at ~20kDa, see figure 6.4), but still higher than expected, it is likely that the predicted highly acidic pI of these proteins affect their behavior upon electrophoresis.

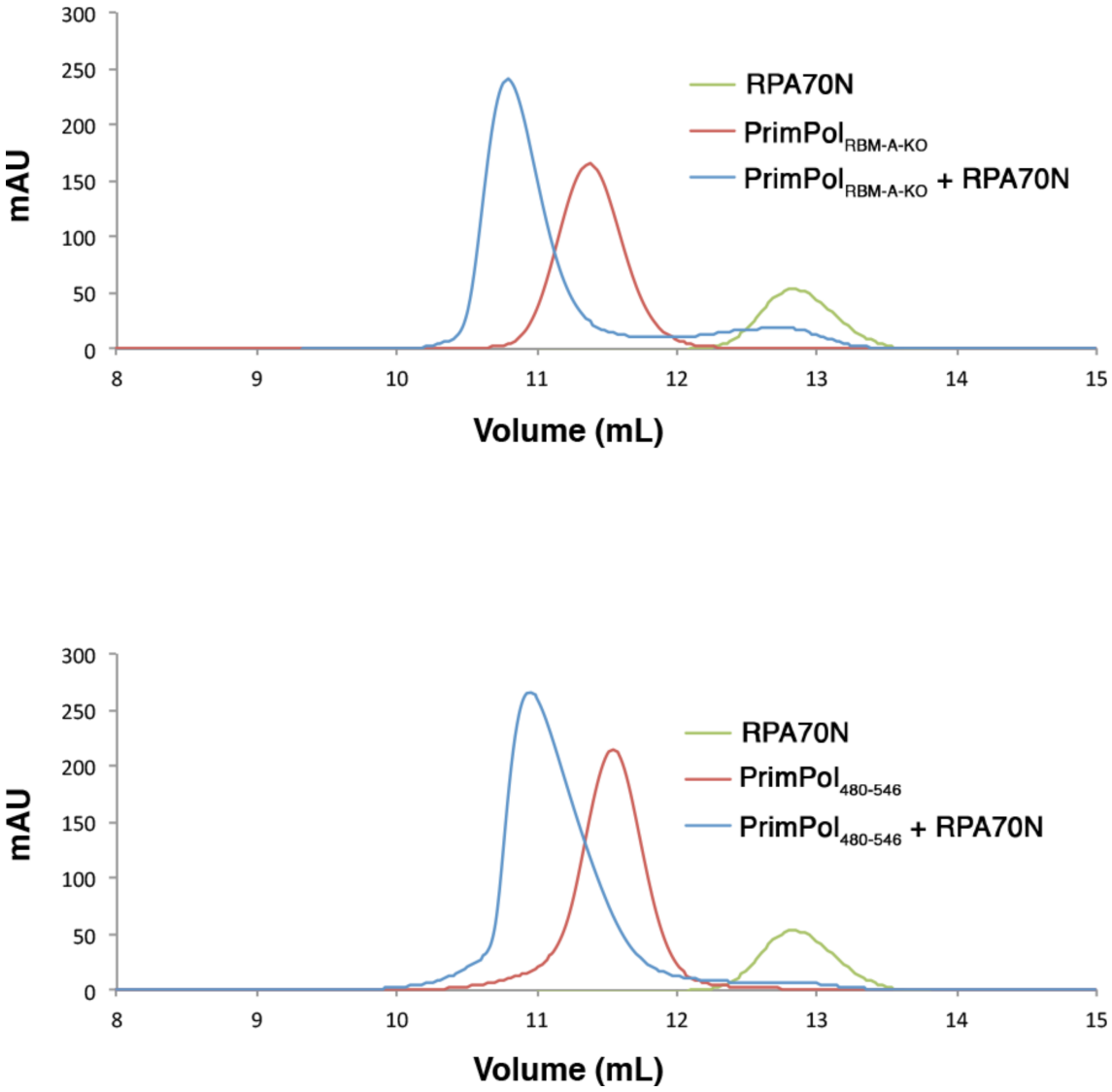


Figure 6.20. PrimPol has two RPA70N binding motifs.
(A) Chromatograph showing the retention volumes of $PrimPol_{480\text{-}560/RBM\text{-}A\text{-}KO}$ (Red), RPA70N (Green) and $PrimPol_{480\text{-}560/RBM\text{-}A\text{-}KO}$ and RPA70N in a 1:1 ratio (Blue). RPA70N has a retention peak volume of 12.86 mL and $PrimPol_{480\text{-}560}$ has a retention peak volume of 11.37 mL. The $PrimPol_{480\text{-}560/RBM\text{-}A\text{-}KO}$-RPA70N mixture chromatograph shows both individual proteins' peaks abrogated and a single peak at a retention volume of 10.87 mL, demonstrating an interaction between these two proteins. **(B)** Chromatograph showing the retention volumes of $PrimPol_{480\text{-}546}$ (Red), RPA70N (Green) and $PrimPol_{480\text{-}546}$ and RPA70N in a 1:1 ratio (Blue). RPA70N has a retention peak volume of 12.86 mL and $PrimPol_{480\text{-}546}$ has a retention peak volume of 11.54 mL. These two proteins also interact, illustrated by a single peak at 10.96 mL.

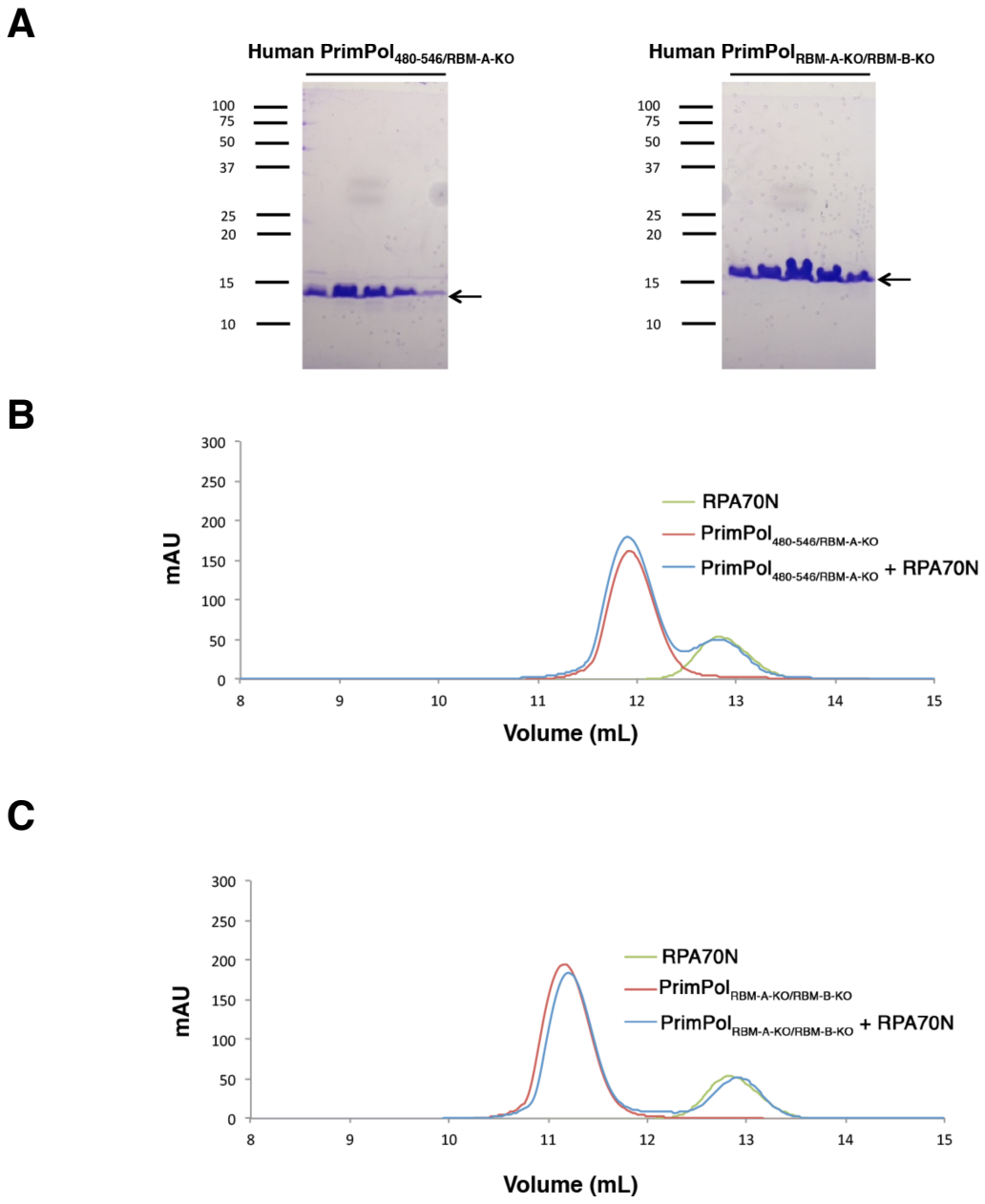


Figure 6.21. RBM-A or RBM-B is required for RPA70N binding.
(A) Gel Filtration of PrimPol$_{480\text{-}546/RBM\text{-}A\text{-}KO}$ and PrimPol $_{RBM\text{-}A\text{-}KO/RBM\text{-}B\text{-}KO}$ RBM-A and RBM-B double knockouts. **(B)** Chromatograph showing the retention volumes of PrimPol$_{480\text{-}546/RBM\text{-}A\text{-}KO}$ (Red), RPA70N (Green) and PrimPol$_{480\text{-}546/RBM\text{-}A\text{-}KO}$ and RPA70N in a 1:1 ratio (Blue). There is no shift in the peaks of the individual proteins, suggesting that there is no interaction between these proteins. **(C)** Chromatograph showing the retention volumes of PrimPol $_{RBM\text{-}A\text{-}KO/RBM\text{-}B\text{-}KO}$ (Red), RPA70N (Green) and PrimPol$_{480\text{-}546}$ and RPA70N in a 1:1 ratio (Blue). Again, there is no shift in either of the individual protein peaks, suggesting that there is no interaction between these proteins. Both these RBM-A and RBM-B double knockouts remove the capacity to bind RPA70N.

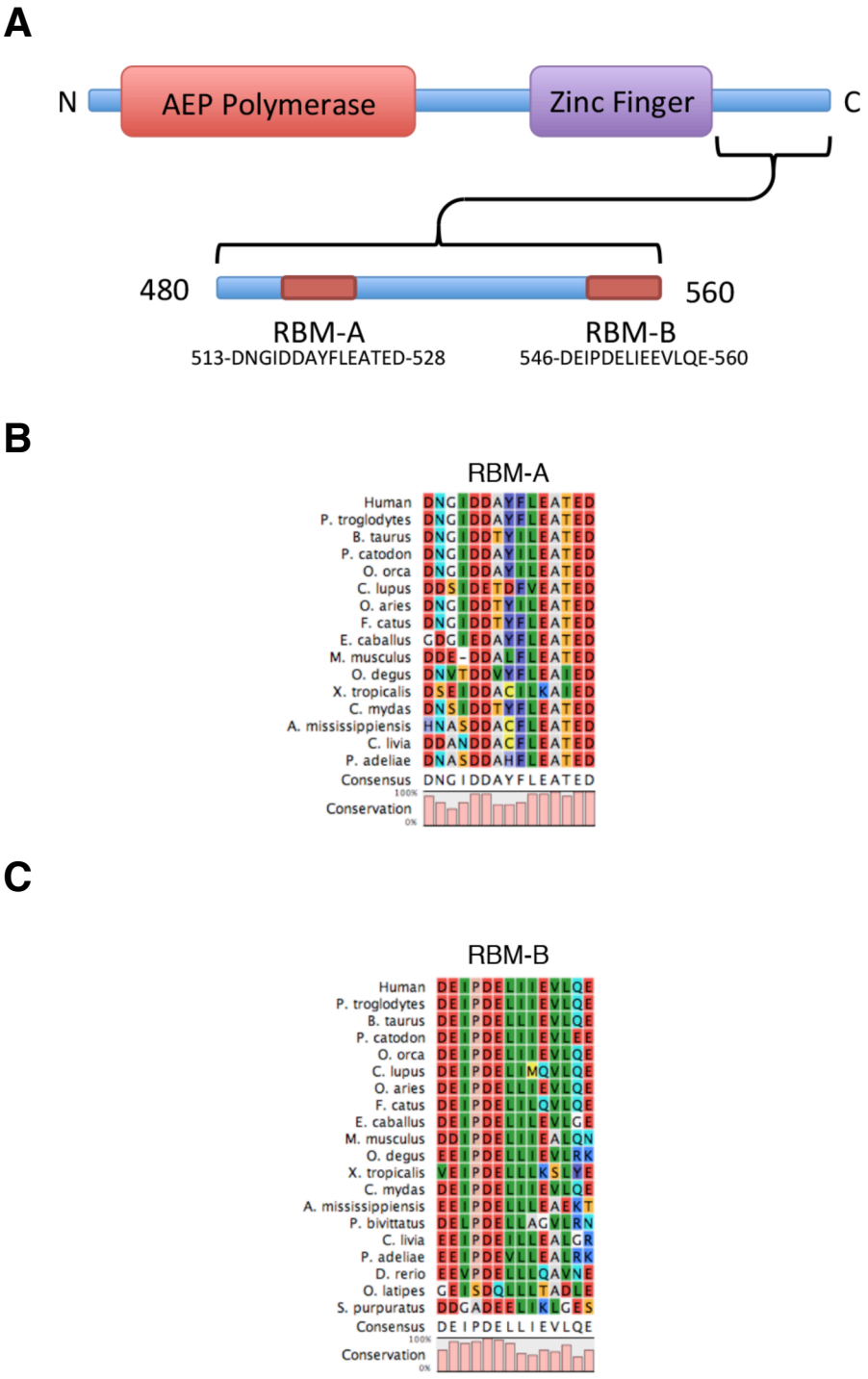


Figure 6.22. RBM-A and RBM-B are conserved motifs.
(A) The domain architecture of PrimPol showing the positions of RBM-A and RBM-B relative to the functional domains of the protein. **(B)** RBM-A is a conserved motif across chordate species. However, in lower species not shown here the motif splits into two separate motifs. **(C)** RBM-B is conserved across higher eukaryotic species.

some PrimPol orthologues have been identified with RBM-A that don't necessarily contain RBM-B, such as the PrimPol orthologue of *Ectocarpus siliculosus*, which contains the sequence NPSIDDGYFLSAGED towards its C-terminus but no apparent homology to RBM-B. It remains difficult to predict which of the two motifs are more biologically important and *in vivo* analysis using human PrimPol variants with RBM-A and RBM-B knocked out is underway in our laboratory.

Bioinformatics analysis using PSI-BLAST reveals that both RBM-A- and RBM-B-like sequences are found in other proteins involved in DNA metabolism. As described, RBM-A was identified due to its homology with other DNA damage response proteins that bind RPA (Section 6.4). We have also identified highly similar sequences found in both Topo I and the Werner helicase (Figure 6.23A). The Werner helicase is a RecQ helicase with crucial roles in double strand break repair, both HR and NHEJ, and base excision repair (Monnat, 2010). RPA is a known interacting partner of the Werner helicase, acting to increase its helicase activity *in vitro* (Brosh et al., 1999; K. M. Doherty et al., 2005). This interaction is mapped to a region on the Werner helicase that has a repeated sequence, both repeats contain the RBM-A-like sequence (K. M. Doherty et al., 2005). RPA70 binds to this repeated sequence but also binds to one individual sequence (K. M. Doherty et al., 2005); the significance of the sequence repeat is unknown and this repeat is not present in the *Xenopus* Werner helicase.

RBM-A also aligns to a sequence in the catalytic subunit (p180, POLA1) of pol α in the fungal species *Pseudozyma antarctica* (Figure 6.23B). A subsequent alignment of this sequence identifies a comparable sequence in the homologous protein in *Saccharomyces cerevisiae* and humans (Figure 6.23B). This acidic sequence is found towards the N-terminus of the protein, amino acids 56-70 in human pol α. The processivity and polymerase activity of pol α is stimulated through interaction with RPA (Braun et al., 1997) and the docking site on RPA has been identified in the region 1-173 of RPA70, i.e. that region containing RPA70N (Taneja et al., 2007). We postulate that these are the most likely residues on pol α to interact with RPA, and do so in a similar manner to PrimPol.

RBM-B of PrimPol strongly resembles a Ctf4-interaction motif (Figure 6.23C) that is also present in the catalytic subunit of pol α the Sld5 subunit of GINS (A. C. Simon et al., 2014). A homotrimer of Ctf4 molecules acts to link pol α to the CMG complex (A. C.

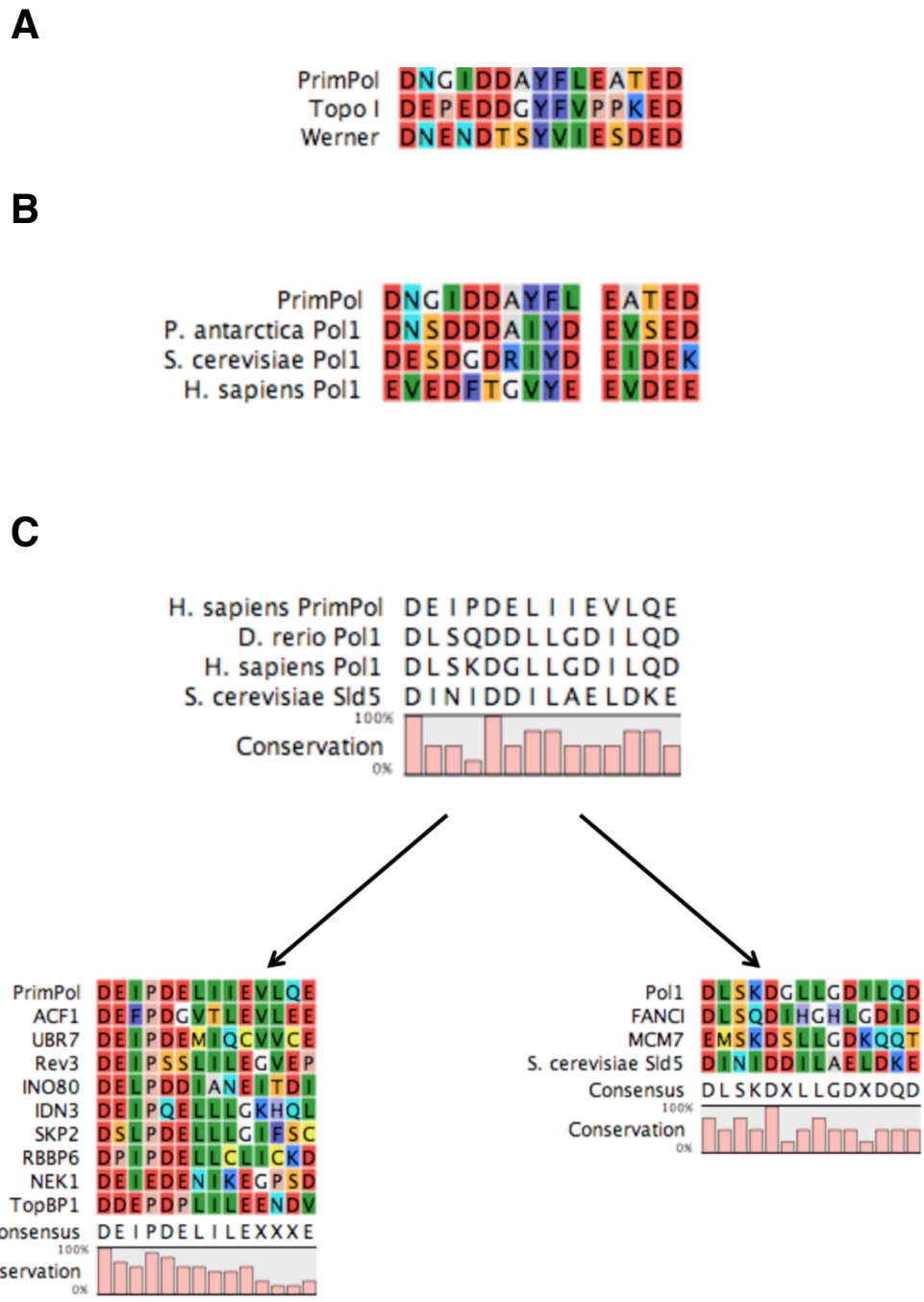


Figure 6.23. RBM-A and RBM-B represent common protein interaction motifs.
(A) RBM-A of PrimPol aligns to two DNA processing proteins. Sequences similar to that of RBM-A are identified in Topoisomerase I and the Werner helicase. **(B)** RBM-A of PrimPol aligns to a short acidic sequence in the catalytic subunit of pol α of *Pseudozyma antarctica*. A comparable sequence is found in the human homologue and it is postulated that this is the RPA-interacting region of pol α. **(C)** PrimPol RBM-B resembles a Ctf4-binding motif. The Ctf4-binding motif that has previously been identified in the catalytic polymerase subunit of pol α (Pol1) and Sld5 of GINS is present in PrimPol at the C-terminus in RBM-B. This region of PrimPol interacts with RPA70N as shown crystallographically and represents a potential multiple protein binding motif. A PSI-BLAST of this motif from human PrimPol and pol α identifies a number of other proteins involved in the metabolism of DNA.

Simon et al., 2014). These motifs also form an α helix and bind Ctf4 in an antiparallel fashion to two helices in Ctf4 that form a basic binding channel (A. C. Simon et al., 2014). It was noted that Ctf4 may be anchored to the CMG complex, it has a higher binding efficiency to Sld5 than pol α (A. C. Simon et al., 2014), and allows dynamic interaction of other replication components as required, allowing efficient replication both under normal conditions and during replication stress. PrimPol is not only active under situations of replicative stress, it is also important for the maintenance of unperturbed replication fork progression ((Bianchi et al., 2013); Section 3.11). Although RPA is essential for PrimPol foci formation during replicative stress (Wan et al., 2013), PrimPol may be brought to the replication fork by binding transiently to Ctf4 as required during unperturbed replication. Further interaction studies are required to validate this hypothesis.

A PSI-BLAST sequence alignment using currently the interaction motifs from PrimPol and pol α identifies range of other proteins that have similar sequences to these motifs, most of which are involved in DNA metabolism (Figure 6.23C). Those with motifs similar to the PrimPol RBM-B motif include ACF1, UBR7, Rev3, INO80, IDN3, RBBP6, NEK1 and TopBP1. Rev3 is the catalytic domain of pol ζ, and could provide a role similar to PrimPol and pol α in this context. INO80 and ACF1 are parts chromatin-remodelling proteins and involved sliding the nucleosomes for DNA replication and DNA repair (A. J. Morrison and Shen, 2009; Poot et al., 2000); INO80 and ACF1 may interact with the Ctf4 complex to slide upstream nucleosomes to allow other Ctf4-associated proteins to carry out their function. Nek1 is a kinase that interacts with Ku and ATR and primes Ku for the loading of other chromatin remodelling proteins, and primes ATR for efficient DNA damage response (S. Liu et al., 2013; Patil et al., 2013). Ctf4 could, therefore, act as the docking partner that brings Nek1 to the site of Ku and ATR. Notably, the aspartate at position 551 of PrimPol that is mainly responsible for the electrostatic interactions with RPA70N is highly conserved throughout these motifs.

Those with motifs similar to the pol α Ctf4-interacting motif include replication proteins MCM7 and DNA repair protein FANCI (Figure 6.23C). MCM7 is another protein of the CMG complex and suggests that the CMG complex could be anchored to the Ctf4 through two separate interactions from its three interaction domains. FANCI forms a complex with FANCD2 is responsible for the repair of interstrand DNA crosslinks (Joo et al., 2011). The identification of this motif raises the possibility that the Ctf4 trimer

acts not only to anchor the CMG complex to pol α but interacts with a multitude of proteins, including PrimPol, to carry out efficient replication.

6.8. Summary and Discussion

Previous work from our laboratory and others have demonstrated that PrimPol interacts with RPA, and that this interaction is essential for foci formation (Guilliam et al., 2015; Wan et al., 2013). The work presented in this chapter defines the interaction between RPA and PrimPol at two binding motifs in the unstructured C-terminus of PrimPol and is the result of a fruitful collaboration between our laboratory and the Chazin laboratory. This interaction and regulation by RPA illustrates a novel regulatory mechanism, distinct from that employed by other TLS polymerases (Lehmann et al., 2007).

We initially demonstrated that the part of PrimPol C-terminal of the zinc finger is poorly structured and reduce the protein's stability, suggesting an unstructured C-terminal tail that is potentially involved in protein-protein interactions. Additionally Wan et al. have shown that the C-terminal 80 amino acids of PrimPol (480-560) are sufficient for foci formation *in vivo* (Wan et al., 2013).

RPA has two predominant protein-interaction domains in the RPA70N OB fold and the winged helix domain in the C-terminus. HSQC NMR spectral shifts following the titration of PrimPol establish that it is the OB fold of RPA70N that binds to PrimPol, not the winged helix domain. Following the removal of the C-terminal amino acids 488-560 this interaction is lost. To verify that the RPA-interaction motif of PrimPol lies in the C-terminus, $PrimPol_{480-560}$ was constructed and shown to directly interact with RPA70N. It was previously postulated that RPA70C was the RPA interacting partner for PrimPol (Wan et al., 2013), but this is disputed by our data and there is no evidence of any other proteins interacting with this portion of RPA.

The studies described here have identified two negatively-charged motifs in the C-terminus of PrimPol that interact with the basic cleft of RPA70N. RPA70N interacts with a number of DNA damage response proteins, including ATRIP, MRE11, RAD9 and p53 and these proteins share a relatively conserved interaction motif, a motif which is also present in PrimPol (Figure 6.8; RBM-A). HSQC NMR spectral shifts through the titration of this protein to ^{15}N-enriched RPA70N demonstrated that this motif is indeed an RPA-interacting motif. However, we also report a co-crystal structure of RPA70N with the

extreme C-terminus of PrimPol (RBM-B). Although a structure of a fusion peptide of RPA70N with p53 has previously been determined (Bochkareva et al., 2005), and a modified, stapled ATRIP peptide has been co-crystallised with RPA70N (Frank et al., 2014), this is the first X-ray crystal structure of an unmodified human protein in the basic binding cleft RPA70N.

This crystal structure reveals a novel binding mode to RPA70N than previously described structures. Previous structures are centred around an interaction in a hydrophobic pocket on RPA70N. Whilst RBM-B has an isoleucine that extends into this hydrophobic pocket and forms hydrophobic interactions, the major contacts are electrostatic interactions towards the N-terminus of the short RBM-B helix, with D551 of RBM-B forming a number of electrostatic interactions at the interface of the two proteins. There are no interactions equivalent to these electrostatic interactions in previous structures as their N-termini extends away from the hydrophobic pocket in the opposite direction than PrimPol's.

PrimPol has two protein binding sites that both bind RPA70N; these binding sites, particularly RBM-B, are likely to bind other protein partners such as Ctf4 to form a complex regulation system for PrimPol. The HSQC spectrum shift of $PrimPol_{480-560}$ was pronounced and distribution was wide when RPA70N was titrated (Figure 6.7), and one potential indication is that there is heterogeneous binding to the binding site. There could, therefore, be exchange events between the two binding motifs at RBM-A and RBM-B. Mutagenesis studies are required to define which of the two motifs is more superior at binding RPA and is responsible for foci formation *in vivo*. It may be that 545-560 is the more relevant binding motif, which is why it is present in the crystal structure, or it could be that 513-527 interacts with RPA *in vivo* but is less stable or somehow precluded in the $PrimPol_{480-560}$ variant that was utilised to obtain the crystal structure.

The data presented here complements the biochemical work carried out in parallel by Dr. Stanislaw Jozwiakowski and Thomas Guilliam from our laboratory. Their recent publication describes the regulatory effects of single stranded binding (SSB) proteins RPA and mtSSB on PrimPol. It was found that these SSB proteins act to constrain both the polymerase and primase activities of PrimPol (Guilliam et al., 2015), thus providing the opposite effect that it has on the high-fidelity replicative polymerases α, δ and γ (Braun et al., 1997; Kenny et al., 1989; Oliveira and Kaguni, 2010; Tsurimoto and

Stillman, 1989). This is not the first example of single-stranded binding proteins acting to restrict the activity of low-fidelity polymerases, with *E. coli* SSB selectively inhibiting the TLS polymerase pol II (Indiani et al., 2013).

Taken together, three mechanisms of regulation for PrimPol can be proposed. The primary regulation mechanism is an intrinsic one; PrimPol only inserts 4 nucleotides in a single binding event (Section 4.2). This prevents the polymerase continuing to polymerise long stretches of DNA, potentially introducing insertion, deletion or point mutations into the genome in doing so.

The other two mechanisms of regulation involve recruitment to the genome as required by protein partners. The first of these is regulation by single-stranded binding proteins RPA and mtSSB (Figure 6.24). RPA and mtSSB stimulate replicative polymerases during unperturbed replication and are displaced by the replication machinery as it replicates in a 5'-3' direction. Upon reaching DNA damage, replicative polymerases stall and idle prior to the site of DNA damage; PCNA becomes monoubiquitinated and the replicative polymerase is relocated from the site of damage. RPA molecules coating the 5' side of the site of damage remain intact and not displaced by the replication machinery. RPA70N recruits PrimPol to carry out TLS past the site of damage, or repriming following the site of damage (Section 4.11; Figure 4.16). RPA prevents PrimPol from excessive replication following the site of damage, as it is inhibitory to PrimPol and PrimPol cannot displace RPA (Guilliam et al., 2015). Following replication past the site of DNA damage by PrimPol, the replicative polymerase can proceed and continue replication.

The second potential regulation mechanism is through tethering to the CMG complex via the Ctf4 trimer (Figure 6.25). The Ctf4 trimer is known to anchor the other primase present in higher eukaryotes, pol α, to the CMG complex (A. C. Simon et al., 2014). The potential interaction with Ctf4 could allow for pol α priming at sites of replication initiation and PrimPol priming following sites of UV damage (Section 4.10). Alternatively, tethering to the CMG complex could be required for as yet undefined PrimPol role in unperturbed replication ((Bianchi et al., 2013); Section 3.11). Future interaction studies with Ctf4 are required to establish if a *bona fide* interaction with PrimPol exists.

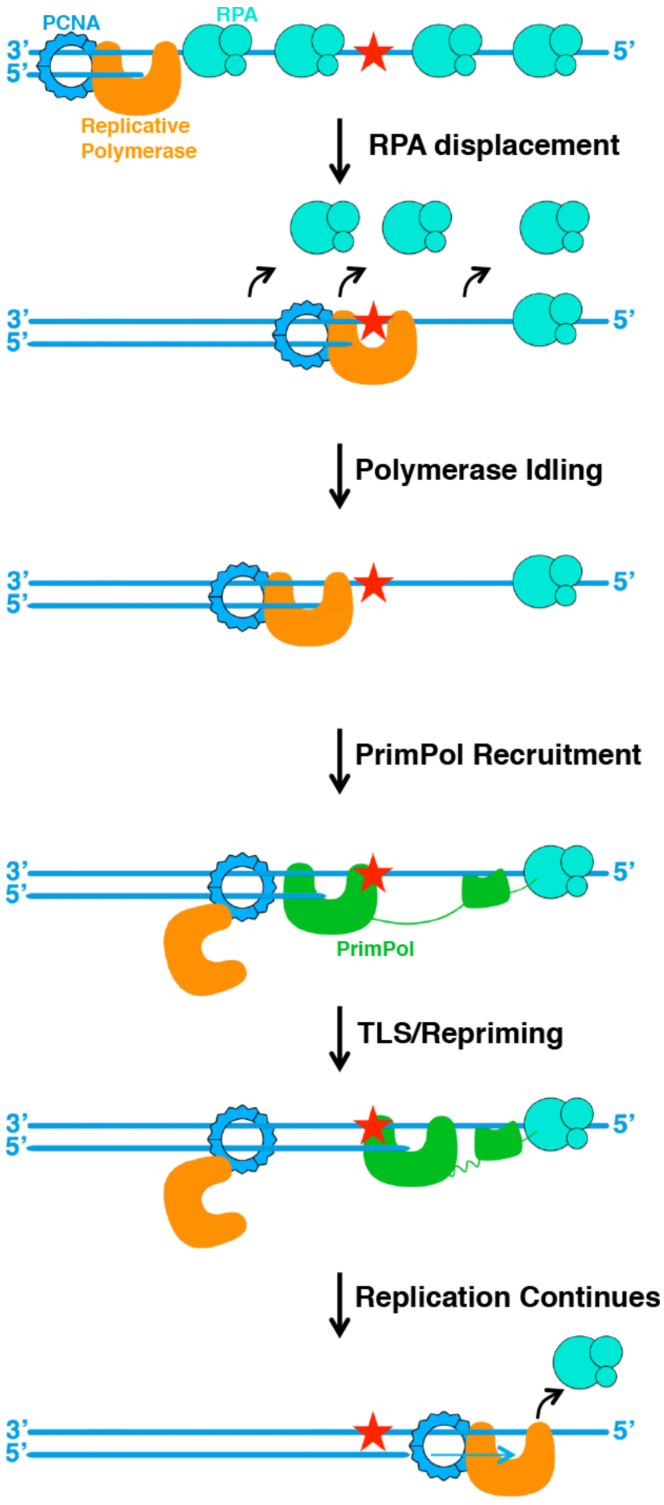


Figure 6.24. Model of regulation of PrimPol by RPA.
The replicative polymerases are proficient at displacing RPA from coated ssDNA as they replicate DNA in a processive manner. As the polymerase reaches a DNA lesion, the polymerase idles as the 3'-5' exonucleolytic activity of the replicative polymerase prevents replication past the dimer. PrimPol binds to the RPA that has not been displaced downstream of the lesion and can extend the nascent DNA chain through the lesion or reprime following the lesion if unable to carry out TLS. The replicative machinery can then continue replication following lesion bypass.

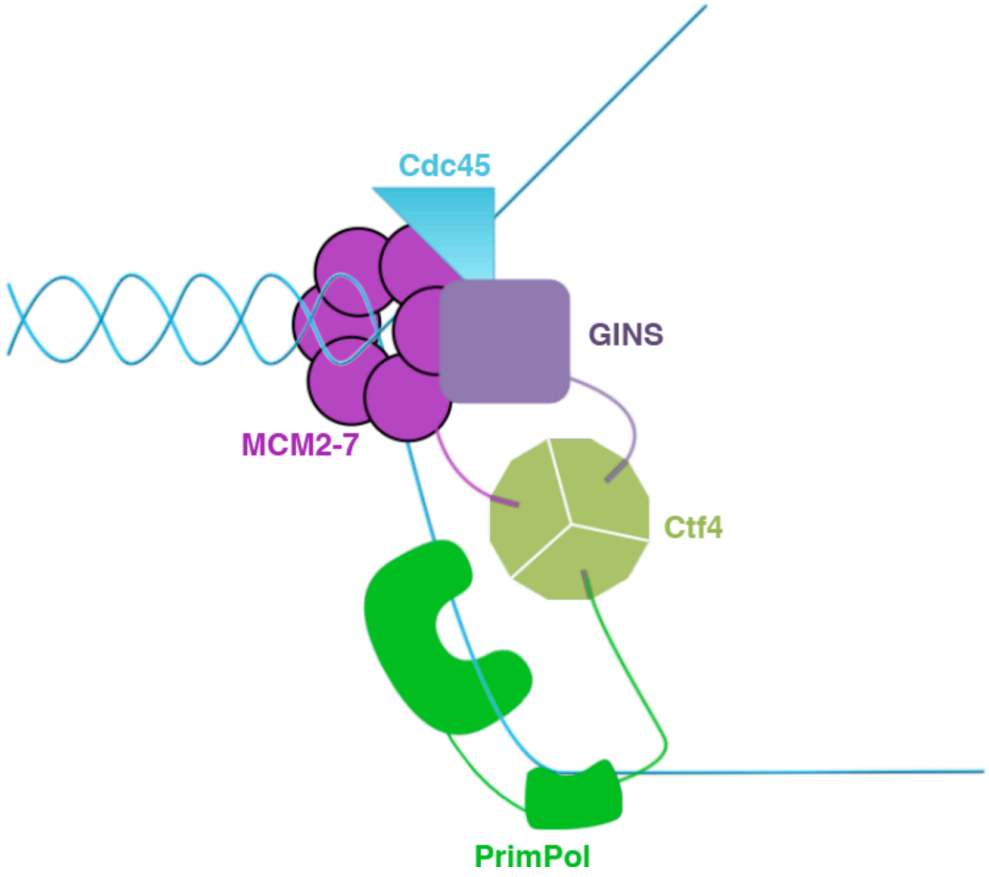


Figure 6.25. Model of regulation of PrimPol by Ctf4.
Ctf4 acts as a trimer and anchors interacting proteins at the replication fork. An interaction between GINS and pol α via Ctf4 has previously been established (Simon et al., 2014). PrimPol and MCM7 also contain potential Ctf4-binding motifs. Ctf4 could act to anchor PrimPol to the CMG helicase complex via Sld5 and MCM7 of the CMG complex to allow PrimPol to reprime following lesions or during PrimPol's currently undefined role in unperturbed replication.

References

Acharya, N., Johnson, R.E., Prakash, S., Prakash, L., 2006. Complex formation with Rev1 enhances the proficiency of Saccharomyces cerevisiae DNA polymerase zeta for mismatch extension and for extension opposite from DNA lesions. Mol. Cell. Biol. 26, 9555–9563.

Admire, A., Shanks, L., Danzl, N., Wang, M., Weier, U., Stevens, W., Hunt, E., Weinert, T., 2006. Cycles of chromosome instability are associated with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. Genes & development 20, 159–173.

Ahel, I., Rass, U., El-Khamisy, S.F., Katyal, S., Clements, P.M., McKinnon, P.J., Caldecott, K.W., West, S.C., 2006. The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. Nature 443, 713–716.

Almeida, K.H., Sobol, R.W., 2007. A unified view of base excision repair: Lesion-dependent protein complexes regulated by post-translational modification. DNA repair 6, 695–711.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Andres, S.N., Vergnes, A., Ristic, D., Wyman, C., Modesti, M., Junop, M., 2012. A human XRCC4–XLF complex bridges DNA. Nucleic Acids Res. 40, 1868–1878.

Aparicio, O.M., Stout, A.M., Bell, S.P., 1999. Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc. Natl. Acad. Sci. USA 96, 9130–9135.

Appleby, T.C., Perry, J.K., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D., Wetmore, D.R., McGrath, M.E., Ray, A.S., Sofia, M.J., Swaminathan, S., Edwards, T.E., 2015. Structural basis for RNA replication by the hepatitis C virus polymerase. Science 347, 771–775.

Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., Hanaoka, F., 2001. Centrosome Protein Centrin 2/Caltractin 1 Is Part of the Xeroderma Pigmentosum Group C Complex That Initiates Global Genome Nucleotide Excision Repair. J. Biol. Chem. 276, 18665–18672.

Aravind, L., Koonin, E.V., 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res. 11, 1365–1374.

Aravind, L., Leipe, D.D., Koonin, E.V., 1998. Toprim-a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Res. 26, 4205–4213.

Arezi, B., Kirk, B.W., Copeland, W.C., Kuchta, R.D., 1999. Interactions of DNA with human DNA primase monitored with photoactivatable cross-linking agents: implications for the role of the p58 subunit. Biochemistry 38, 12899–12907.

Augustin, M.A., Huber, R., Kaiser, J.T., 2001. Crystal structure of a DNA-dependent RNA polymerase (DNA primase). Nat. Struct. Biol. 8, 57–61.

Avery, O.T., Macleod, C.M., McCarty, M., 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type III. J. Exp. Med. 79, 137–158.

Azvolinsky, A., Giresi, P.G., Lieb, J.D., Zakian, V.A., 2009. Highly Transcribed RNA Polymerase II Genes Are Impediments to Replication Fork Progression in Saccharomyces cerevisiae. Mol. Cell 34, 722–734.

Barnes, D.E., Johnston, L.H., Kodama, K., Tomkinson, A.E., Lasko, D.D., Lindahl, T.,

1990. Human DNA ligase I cDNA: cloning and functional expression in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 87, 6679–6683.
Bartlett, E.J., Brissett, N.C., Doherty, A.J., 2013. Ribonucleolytic resection is required for repair of strand displaced nonhomologous end-joining intermediates. Proc. Natl. Acad. Sci. USA 110, E1984–91.
Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W., 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. D Biol. Crystallogr. 67, 271–281.
Baxter, J., Diffley, J.F.X., 2008. Topoisomerase II Inactivation Prevents the Completion of DNA Replication in Budding Yeast. Mol. Cell 30, 790–802.
Beck, K., Lipps, G., 2007. Properties of an unusual DNA primase from an archaeal plasmid. Nucleic Acids Res. 35, 5635–5645.
Beese, L.S., Derbyshire, V., Steitz, T.A., 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. Science 260, 352–355.
Bell, S.P., Stillman, B., 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature 357, 128–134.
Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2009. GenBank. Nucleic Acids Res. 37, D26–31.
Bernstein, N.K., Williams, R.S., Rakovszky, M.L., Cui, D., 2005. The Molecular Architecture of the Mammalian DNA Repair Enzyme, Polynucleotide Kinase. Mol. Cell 17, 657–670.
Bertram, J.S., 2000. The molecular biology of cancer. Molecular aspects of medicine 21, 167–223.
Bessette, P.H., Aslund, F., Beckwith, J., Georgiou, G., 1999. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. Proc. Natl. Acad. Sci. USA 96, 13703–13708.
Bianchi, J., Rudd, S.G., Jozwiakowski, S.K., Bailey, L.J., Soura, V., Taylor, E., Stevanovic, I., Green, A.J., Stacker, T.H., Lindsay, H.D., Doherty, A.J., 2013. Eukaryotic PrimPol bypasses UV photoproducts during chromosomal DNA replication. Mol. Cell 52, 566–573.
Biertümpfel, C., Zhao, Y., Kondo, Y., Ramón-Maiques, S., Gregory, M., Lee, J.Y., Masutani, C., Lehmann, A.R., Hanaoka, F., Yang, W., 2010. Structure and mechanism of human DNA polymerase η. Nature 465, 1044–1048.
Biswas, N., Weller, S.K., 1999. A mutation in the C-terminal putative Zn2+ finger motif of UL52 severely affects the biochemical activities of the HSV-1 helicase-primase subcomplex. J. Biol. Chem. 274, 8068–8076.
Blier, P.R., Griffith, A.J., Craft, J., Hardin, J.A., 1993. Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. J. Biol. Chem. 268, 7594–7601.
Blow, J.J., Ge, X.Q., Jackson, D.A., 2011. How dormant origins promote complete genome replication. Trends in Biochemical Sciences 36, 405–414.
Bochkarev, A., Bochkareva, E., 2004. From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. Current Opinion in Structural Biology 14, 36–42.
Bochkareva, E., Kaustov, L., Ayed, A., Yi, G.-S., Lu, Y., Pineda-Lucena, A., Liao, J.C.C., Okorokov, A.L., Milner, J., Arrowsmith, C.H., Bochkarev, A., 2005. Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. Proc. Natl. Acad. Sci. USA 102, 15412–15417.
Bochkareva, E., Korolev, S., Lees-Miller, S.P., Bochkarev, A., 2002. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. The EMBO Journal 21, 1855–1863.
Bocquier, A.A., Liu, L., Cann, I.K.O., Komori, K., Kohda, D., Ishino, Y., 2001. Archaeal

primase: bridging the gap between RNA and DNA polymerases. Current Biology 11, 452–456.
Bouché, J.P., Zechel, K., Kornberg, A., 1975. dnaG gene product, a rifampicin-resistant RNA polymerase, initiates the conversion of a single-stranded coliphage DNA to its duplex replicative form. J. Biol. Chem. 250, 5995–6001.
Boudsocq, F., Kokoska, R.J., Plosky, B.S., Vaisman, A., Ling, H., Kunkel, T.A., Yang, W., Woodgate, R., 2004. Investigating the Role of the Little Finger Domain of Y-family DNA Polymerases in Low Fidelity Synthesis and Translesion Replication. J. Biol. Chem. 279, 32932–32940.
Braithwaite, D.K., Ito, J., 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. 21, 787–802.
Bratic, I., Hench, J., Henriksson, J., Antebi, A., Burglin, T.R., Trifunovic, A., 2009. Mitochondrial DNA level, but not active replicase, is essential for Caenorhabditis elegans development. Nucleic Acids Res. 37, 1817–1828.
Braun, K.A., Lao, Y., He, Z., Ingles, C.J., Wold, M.S., 1997. Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase alpha by multiple mechanisms. Biochemistry 36, 8443–8454.
Brautigam, C.A., Steitz, T.A., 1998. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. Current Opinion in Structural Biology 8, 54–63.
Brewer, B.J., Fangman, W.L., 1988. A replication fork barrier at the 3′ end of yeast ribosomal RNA genes. Cell 55, 637–643.
Brissett, N.C., Martin, M.J., Pitcher, R.S., Bianchi, J., Juarez, R., Green, A.J., Fox, G.C., Blanco, L., Doherty, A.J., 2011. Structure of a preternary complex involving a prokaryotic NHEJ DNA polymerase. Mol. Cell 41, 221–231.
Brissett, N.C., Pitcher, R.S., Juarez, R., Picher, A.J., Green, A.J., Dafforn, T.R., Fox, G.C., Blanco, L., Doherty, A.J., 2007. Structure of a NHEJ polymerase-mediated DNA synaptic complex. Science 318, 456–459.
Brooks, M., Dumas, L.B., 1989. DNA primase isolated from the yeast DNA primase-DNA polymerase complex. Immunoaffinity purification and analysis of RNA primer synthesis. J. Biol. Chem. 264, 3602–3610.
Brosh, R.M., Orren, D.K., Nehlin, J.O., Ravn, P.H., Kenny, M.K., Machwe, A., Bohr, V.A., 1999. Functional and physical interaction between WRN helicase and human replication protein A. J. Biol. Chem. 274, 18341–18350.
Brown, J.A., Suo, Z., 2011. Unlocking the Sugar “Steric Gate” of DNA Polymerases. Biochemistry 50, 1135–1142.
Brutlag, D., Schekman, R., Kornberg, A., 1971. A possible role for RNA polymerase in the initiation of M13 DNA synthesis. Proc. Natl. Acad. Sci. USA 68, 2826–2829.
Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., Jones, D.T., 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 41, W349–57.
Burgers, P.M.J., 1998. Eukaryotic DNA polymerases in DNA replication and DNA repair. Chromosoma 107, 218–227.
Cairns, J., 1966. Autoradiography of HeLa cell DNA. J. Mol. Biol. 15, 372–373.
Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E., 2003. Failure to Produce Mitochondrial DNA Results in Embryonic Lethality in Rnaseh1 Null Mice. Mol. Cell 11, 807–815.
Champoux, J.J., 2001. DNA topoisomerases: structure, function, and mechanism. Annual Review of Biochemistry 70, 369–413.
Chan, S.S.L., Copeland, W.C., 2009. DNA polymerase gamma and mitochondrial disease: understanding the consequence of POLG mutations. Biochim. Biophys.

Acta 1787, 312–319.
Chang, D.D., Clayton, D.A., 1985. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. USA 82, 351–355.
Chang, D.D., Hauswirth, W.W., Clayton, D.A., 1985. Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. The EMBO Journal 4, 1559.
Chargaff, E., Lipshitz, R., Green, C., 1952. Composition of the desoxypentose nucleic acids of four genera of sea-urchin. J. Biol. Chem. 195, 155–160.
Chattopadhyay, S., Bielinsky, A.-K., 2007. Human Mcm10 regulates the catalytic subunit of DNA polymerase-alpha and prevents DNA damage during replication. Mol. Biol. Cell 18, 4085–4095.
Chen, Y., Carrington-Lawrence, S.D., Bai, P., Weller, S.K., 2005. Mutations in the putative zinc-binding motif of UL52 demonstrate a complex interdependence between the UL5 and UL52 subunits of the human herpes simplex virus type 1 helicase/primase complex. J. Virol. 79, 9088–9096.
Chen, Y., Livingston, C.M., Carrington-Lawrence, S.D., Bai, P., Weller, S.K., 2007. A Mutation in the Human Herpes Simplex Virus Type 1 UL52 Zinc Finger Motif Results in Defective Primase Activity but Can Recruit Viral Polymerase and Support Viral Replication Efficiently. J. Virol. 81, 8742–8751.
CLCBio, 2014. CLC Sequence Viewer 7.0.
Coin, F., Oksenych, V., Mocquet, V., Groh, S., Blattner, C., 2008. Nucleotide Excision Repair Driven by the Dissociation of CAK from TFIIH. Mol. Cell 31, 9–20.
Collado, M., Gil, J., Efeyan, A., Guerra, C., Schuhmacher, A.J., Barradas, M., Benguría, A., Zaballos, A., Flores, J.M., Barbacid, M., Beach, D., Serrano, M., 2005. Tumour biology: Senescence in premalignant tumours. Nat. Cell Biol. 436, 642–642.
Conaway, R.C., Lehman, I.R., 1982a. Synthesis by the DNA primase of Drosophila melanogaster of a primer with a unique chain length. Proc. Natl. Acad. Sci. USA 79, 4585–4588.
Conaway, R.C., Lehman, I.R., 1982b. A DNA primase activity associated with DNA polymerase alpha from Drosophila melanogaster embryos. Proc. Natl. Acad. Sci. USA 79, 2523–2527.
Connolly, B.A., Fogg, M.J., Shuttleworth, G., Wilson, B.T., 2003. Uracil recognition by archaeal family B DNA polymerases. Biochemical Society Transactions 31, 699–702.
Copeland, W.C., Tan, X., 1995. Active Site Mapping of the Catalytic Mouse Primase Subunit by Alanine Scanning Mutagenesis. J. Biol. Chem. 270, 3905–3913.
Copeland, W.C., Wang, T.S., 1993. Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. J. Biol. Chem. 268, 26179–26189.
Cori, C.F., Cori, G.T., Green, A.A., 1943. Crystalline Muscle Phosphorylase. J. Biol. Chem. 151, 21–38.
Corn, J.E., Pease, P.J., Hura, G.L., Berger, J.M., 2005. Crosstalk between Primase Subunits Can Act to Regulate Primer Synthesis in trans. Mol. Cell 20, 391–401.
Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P.A., 2007. Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. DNA repair 6, 712–722.
Crick, F., 1970. Central dogma of molecular biology. Nature 227, 561–563.
Crick, F.H., Barnett, L., Brenner, S., Watts-Tobin, R.J., 1961. General nature of the genetic code for proteins. Nature 192, 1227–1232.
Crute, J.J., Lehman, I.R., 1991. Herpes simplex virus-1 helicase-primase. Physical and catalytic properties. J. Biol. Chem. 266, 4484–4488.

Dairaghi, D.J., Shadel, G.S., Clayton, D.A., 1995. Addition of a 29 Residue Carboxyl-terminal Tail Converts a Simple HMG Box-containing Protein into a Transcriptional Activator. J. Mol. Biol. 249, 11–28.

De Falco, M., Fusco, A., De Felice, M., Mosè, R., Pisani, F.M., 2004. The DNA primase of Sulfolobus solfataricus is activated by substrates containing a thymine-rich bubble and has a 3'-terminal nucleotidyl-transferase activity. Nucleic Acids Res. 32, 5223–5230.

De Silva, F.S., Paran, N., Moss, B., 2009. Products and substrate/template usage of vaccinia virus DNA primase. Virology 383, 136–141.

Dean, C.J., Ormerod, M.G., Serianni, R.W., Alexander, P., 1969. DNA strand breakage in cells irradiated with x-rays. Nature 222, 1042–1044.

Deininger, M.W.N., Goldman, J.M., Melo, J.V., 2000. The molecular biology of chronic myeloid leukemia. Blood 96, 3343–3356.

Della, M., Palmbos, P.L., Tseng, H.-M., Tonkin, L.M., Daley, J.M., Topper, L.M., Pitcher, R.S., Tomkinson, A.E., Wilson, T.E., Doherty, A.J., 2004. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. Science 306, 683–685.

Desjardins, P., Frost, E., Morais, R., 1985. Ethidium bromide-induced loss of mitochondrial DNA from primary chicken embryo fibroblasts. Mol. Cell. Biol. 5, 1163–1169.

Desogus, G., Onesti, S., Brick, P., Rossi, M., Pisani, F.M., 1999. Identification and characterization of a DNA primase from the hyperthermophilic archaeon Methanococcus jannaschii. Nucleic Acids Res. 27, 4444–4450.

Devasagayam, T.P.A., Steenken, S., Obendorf, M.S.W., Schulz, W.A., Sies, H., 1991. Formation of 8-hydroxy(deoxy)guanosine and generation of strand breaks at guanine residues in DNA by singlet oxygen. Biochemistry 30, 6283–6289.

Dianova, I.I., Sleeth, K.M., Allinson, S.L., Parsons, J.L., Breslin, C., Caldecott, K.W., Dianov, G.L., 2004. XRCC1–DNA polymerase β interaction is required for efficient base excision repair. Nucleic Acids Res. 32, 2550–2555.

Dizdaroglu, M., 1985. Formation of 8-hydroxyguanine moiety in deoxyribonucleic acid on .gamma.-irradiation in aqueous solution. Biochemistry 24, 4476–4481.

Doherty, K.M., Sommers, J.A., Gray, M.D., Lee, J.W., Kobbe, von, C., Thoma, N.H., Kureekattil, R.P., Kenny, M.K., Brosh, R.M., 2005. Physical and functional mapping of the replication protein a interaction domain of the werner and bloom syndrome helicases. J. Biol. Chem. 280, 29494–29505.

Donigan, K.A., McLenigan, M.P., Yang, W., Goodman, M.F., Woodgate, R., 2014. The steric gate of DNA polymerase ι regulates ribonucleotide incorporation and deoxyribonucleotide fidelity. J. Biol. Chem. 289, 9136–9145.

Dornreiter, I., Erdile, L.F., Gilbert, I.U., Winkler, von, D., Kelly, T.J., Fanning, E., 1992. Interaction of DNA polymerase alpha-primase with cellular replication protein A and SV40 T antigen. The EMBO Journal 11, 769–776.

Drabløs, F., Feyzi, E., Aas, P.A., Vaagbø, C.B., Kavli, B., Bratlie, M.S., Peña-Diaz, J., Otterlei, M., Slupphaug, G., Krokan, H.E., 2004. Alkylation damage in DNA and RNA--repair mechanisms and medical significance. DNA repair 3, 1389–1407.

Duncan, B.K., Miller, J.H., 1980. Mutagenic deamination of cytosine residues in DNA. Nature 287, 560–561.

Duxin, J.P., Dao, B., Martinsson, P., Rajala, N., Guittat, L., Campbell, J.L., Spelbrink, J.N., Stewart, S.A., 2009. Human Dna2 Is a Nuclear and Mitochondrial DNA Maintenance Protein. Mol. Cell. Biol. 29, 4274–4282.

Dzantiev, L., Constantin, N., Genschel, J., Iyer, R.R., Burgers, P.M., Modrich, P., 2004. A Defined Human System That Supports Bidirectional Mismatch-Provoked Excision. Mol. Cell 15, 31–41.

Edgell, D.R., Doolittle, W.F., 1997. Archaea and the origin(s) of DNA replication proteins. Cell 89, 995–998.

Edmunds, C.E., Simpson, L.J., Sale, J.E., 2008. PCNA ubiquitination and REV1 define temporally distinct mechanisms for controlling translesion synthesis in the avian cell line DT40. Mol. Cell 30, 519–529.

Engelke, D.R., Krikos, A., Bruck, M.E., Ginsburg, D., 1990. Purification of Thermus aquaticus DNA polymerase expressed in Escherichia coli. Analytical Biochemistry 191, 396–400.

Fachinetti, D., Bermejo, R., Cocito, A., Minardi, S., Katou, Y., 2010. Replication Termination at Eukaryotic Chromosomes Is Mediated by Top2 and Occurs at Genomic Loci Containing Pausing Elements. Mol. Cell.

Fairman, M., Prelich, G., Tsurimoto, T., Stillman, B., 1988. Identification of cellular components required for SV40 DNA replication in vitro. Biochim. Biophys. Acta 951, 382–387.

Fairman, M.P., Stillman, B., 1988. Cellular factors required for multiple stages of SV40 DNA replication in vitro. The EMBO Journal 7, 1211–1218.

Fan, L., Fuss, J.O., Cheng, Q.J., Arvai, A.S., Hammel, M., Roberts, V.A., Cooper, P.K., Tainer, J.A., 2008. XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. Cell 133, 789–800.

Fan, L., Sanschagrin, P.C., Kaguni, L.S., Kuhn, L.A., 1999. The accessory subunit of mtDNA polymerase shares structural homology with aminoacyl-tRNA synthetases: Implications for a dual role as a primer recognition factor and processivity clamp. Proc. Natl. Acad. Sci. USA 96, 9527–9532.

Feldkamp, M.D., Frank, A.O., Kennedy, J.P., Patrone, J.D., Vangamudi, B., Waterson, A.G., Fesik, S.W., Chazin, W.J., 2013. Surface Reengineering of RPA70N Enables Cocrystallization with an Inhibitor of the Replication Protein A Interaction Motif of ATR Interacting Protein. Biochemistry 52, 6515–6524.

Feng, L., Chen, J., 2012. The E3 ligase RNF8 regulates KU80 removal and NHEJ repair. Nature Structural & Molecular Biology 19, 201–206.

Fiala, K.A., Suo, Z., 2004. Pre-steady-state kinetic studies of the fidelity of Sulfolobus solfataricus P2 DNA polymerase IV. Biochemistry 43, 2106–2115.

Fisher, R.P., Topper, J.N., Clayton, D.A., 1987. Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. Cell 50, 247–258.

Fogg, M.J., Pearl, L.H., Connolly, B.A., 2002. Structural basis for uracil recognition by archaeal family B DNA polymerases. Nat. Struct. Biol. 9, 922–927.

Fortini, P., Pascucci, B., Parlanti, E., Sobol, R.W., Wilson, S.H., Dogliotti, E., 1998. Different DNA Polymerases Are Involved in the Short- and Long-Patch Base Excision Repair in Mammalian Cells. Biochemistry 37, 3575–3580.

Fousteri, M., Vermeulen, W., van Zeeland, A.A., Mullenders, L.H.F., 2006. Cockayne Syndrome A and B Proteins Differentially Regulate Recruitment of Chromatin Remodeling and Repair Factors to Stalled RNA Polymerase II In Vivo. Mol. Cell 23, 471–482.

Frank, A.O., Vangamudi, B., Feldkamp, M.D., Souza-Fagundes, E.M., Luzwick, J.W., Cortez, D., Olejniczak, E.T., Waterson, A.G., Rossanese, O.W., Chazin, W.J., Fesik, S.W., 2014. Discovery of a potent stapled helix peptide that binds to the 70N domain of replication protein A. J. Med. Chem. 57, 2455–2461.

Frick, D.N., Richardson, C.C., 2001. DNA primases. Annual Review of Biochemistry 70, 39–80.

Friedberg, E.C., Lehmann, A.R., Fuchs, R.P.P., 2005. Trading places: how do DNA polymerases switch during translesion DNA synthesis? Mol. Cell 18, 499–505.

Friedberg, E.C., Walker, G.C., Siede, W., 1995. DNA Repair and Mutagenesis. ASM

Press.
Fu, Y.V., Yardimci, H., Long, D.T., Ho, T.V., Guainazzi, A., Bermudez, V.P., Hurwitz, J., van Oijen, A., Schärer, O.D., Walter, J.C., 2011. Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. Cell 146, 931–941.
Futami, K., Shimamoto, A., Furuichi, Y., 2007. Mitochondrial and Nuclear Localization of Human Pif1 Helicase. Biol. Pharm. Bull. 30, 1685–1692.
Gadsden, M.H., McIntosh, E.M., Game, J.C., Wilson, P.J., Haynes, R.H., 1993. dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisiae. The EMBO Journal 12, 4425–4431.
Galal, W.C., Pan, M., Giulian, G., Yuan, W., Li, S., Edwards, J.L., Marino, J.P., Kelman, Z., Hurwitz, J., 2012a. Formation of dAMP-glycerol and dAMP-Tris Derivatives by Thermococcus kodakaraensis DNA Primase. J. Biol. Chem. 287, 16220–16229.
Galal, W.C., Pan, M., Kelman, Z., Hurwitz, J., 2012b. Characterization of DNA Primase Complex Isolated from the Archaeon, Thermococcus kodakaraensis. J. Biol. Chem. 287, 16209–16219.
Game, J.C., Mortimer, R.K., 1974. A genetic study of x-ray sensitive mutants in yeast. Mutation Research 24, 281–292.
Garcia-Gomez, S., Reyes, A., Martínez-Jiménez, M.I., Chocrón, E.S., Mourón, S., Terrados, G., Powell, C., Salido, E., Mendez, J., Holt, I.J., Blanco, L., 2013. PrimPol, an Archaic Primase/Polymerase Operating in Human Cells. Mol. Cell 52, 541–553.
Gary, R., 1999. Proliferating Cell Nuclear Antigen Facilitates Excision in Long-patch Base Excision Repair. J. Biol. Chem. 274, 4354–4363.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker, J.M. (Ed.), The Proteomics Protocols Handbook. Humana Press, pp. 571–607.
Geacintov, N.E., Hibshoosh, H., Ibanez, V., Benjamin, M.J., 1984. Mechanisms of reaction of benzo(a)pyrene-7,8-diol-9,10-epoxide with DNA in aqueous solutions. Biophysical Chemistry 20, 121–133.
Ghosh, A., Joy, A., Schuster, G.B., Douki, T., Cadet, J., 2008. Selective one-electron oxidation of duplex DNA oligomers: reaction at thymines. Org. Biomol. Chem. 6, 916–928.
Gill, S., Krupovic, M., Desnoues, N., Beguin, P., Sezonov, G., Forterre, P., 2014. A highly divergent archaeo-eukaryotic primase from the Thermococcus nautilus plasmid, pTN2. Nucleic Acids Res. 42, 3707–3719.
Gong, C., Bongiorno, P., Martins, A., Stephanou, N.C., Zhu, H., Shuman, S., Glickman, M.S., 2005. Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nature Structural & Molecular Biology 12, 304–312.
Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M., 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. 17, 4713–4730.
Gradia, S., 2000. The Role of Mismatched Nucleotides in Activating the hMSH2-hMSH6 Molecular Switch. J. Biol. Chem. 275, 3922–3930.
Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., Lieber, M.R., 1997. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature 388, 492–495.
Graziewicz, M.A., Bienstock, R.J., Copeland, W.C., 2007. The DNA polymerase gamma Y955C disease variant associated with PEO and parkinsonism mediates the incorporation and translesion synthesis opposite 7,8-dihydro-8-oxo-2'-deoxyguanosine. Hum. Mol. Genet. 16, 2729–2739.

Greenfeder, S.A., Newlon, C.S., 1992a. A replication map of a 61-kb circular derivative of Saccharomyces cerevisiae chromosome III. Mol. Biol. Cell 3, 999–1013.
Greenfeder, S.A., Newlon, C.S., 1992b. Replication forks pause at yeast centromeres. Mol. Cell. Biol. 12, 4056–4066.
Greider, C.W., Blackburn, E.H., 1989. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature 337, 331–337.
Griffith, F., 1928. The Significance of Pneumococcal Types. J. Hyg. 27, 113–159.
Grundy, G.J., Rulten, S.L., Zeng, Z., Arribas-Bosacoma, R., Iles, N., Manley, K., Oliver, A., Caldecott, K.W., 2013. APLF promotes the assembly and activity of non-homologous end joining protein complexes. The EMBO Journal 32, 112–125.
Guilliam, T.A., Jozwiakowski, S.K., Ehlinger, A., Barnes, R.P., Rudd, S.G., Bailey, L.J., Skehel, J.M., Eckert, K.A., Chazin, W.J., Doherty, A.J., 2015. Human PrimPol is a highly error-prone polymerase regulated by single-stranded DNA binding proteins. Nucleic Acids Res. 43, 1056–1068.
Halgasova, N., Mesarosova, I., Bukovska, G., 2012. Identification of a bifunctional primase-polymerase domain of corynephage BFK20 replication protein gp43. Virus Res. 163, 454–460.
Hall, D.B., Holmlin, R.E., Barton, J.K., 1996. Oxidative DNA damage through long-range electron transfer. Nature 382, 731–735.
Halliwell, B., Aruoma, O.I., 1991. DNA damage by oxygen-derived species Its mechanism and measurement in mammalian systems. FEBS Lett. 281, 9–19.
Hammel, M., Rey, M., Yu, Y., Mani, R.S., Classen, S., Liu, M., Pique, M.E., Fang, S., Mahaney, B.L., Weinfeld, M., Schriemer, D.C., Lees-Miller, S.P., Tainer, J.A., 2011. XRCC4 Protein Interactions with XRCC4-like Factor (XLF) Create an Extended Grooved Scaffold for DNA Ligation and Double Strand Break Repair. J. Biol. Chem. 289, 32638–32650.
Hanada, K., Budzowska, M., Davies, S.L., van Drunen, E., Onizawa, H., Beverloo, H.B., Maas, A., Essers, J., Hickson, I.D., Kanaar, R., 2007. The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. Nature Structural & Molecular Biology 14, 1096–1104.
Hanada, K., Budzowska, M., Modesti, M., Maas, A., Wyman, C., Essers, J., Kanaar, R., 2006. The structure-specific endonuclease Mus81–Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. The EMBO Journal 25, 4921–4932.
Hance, N., Ekstrand, M.I., Trifunovic, A., 2005. Mitochondrial DNA polymerase gamma is essential for mammalian embryogenesis. Hum. Mol. Genet. 14, 1775–1783.
Haracska, L., Yu, S.L., Johnson, R.E., Prakash, L., Prakash, S., 2000. Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta. Nat. Genet. 25, 458–461.
Hayakawa, H., Taketomi, A., Sakumi, K., Kuwano, M., Sekiguchi, M., 1995. Generation and elimination of 8-oxo-7,8-dihydro-2“-deoxyguanosine 5-”triphosphate, a mutagenic substrate for DNA synthesis, in human cells. Biochemistry 34, 89–95.
Hays, H., Berdis, A.J., 2002. Manganese substantially alters the dynamics of translesion DNA synthesis. Biochemistry 41, 4771–4778.
Hazra, T.K., Das, A., Das, S., Choudhury, S., Kow, Y.W., Roy, R., 2007. Oxidative DNA damage repair in mammalian cells: a new perspective. DNA repair 6, 470–480.
Helleday, T., Lo, J., van Gent, D.C., Engelward, B.P., 2007. DNA double-strand break repair: from mechanistic understanding to cancer treatment. DNA repair 6, 923–935.
Hershey, A.D., Chase, M., 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. 36, 39–56.

Hess, M.T., Gunz, D., Luneva, N., Geacintov, N.E., Naegeli, H., 1997. Base pair conformation-dependent excision of benzo[a]pyrene diol epoxide-guanine adducts by human nucleotide excision repair enzymes. Mol. Cell. Biol. 17, 7069–7076.

Hines, J.C., Ray, D.S., 2010. A mitochondrial DNA primase is essential for cell growth and kinetoplast DNA replication in Trypanosoma brucei. Mol. Cell. Biol. 30, 1319–1328.

Hines, J.C., Ray, D.S., 2011. A second mitochondrial DNA primase is essential for cell growth and kinetoplast minicircle DNA replication in Trypanosoma brucei. Eukaryotic Cell 10, 445–454.

Hitomi, K., Iwai, S., Tainer, J.A., 2007. The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. DNA repair 6, 410–428.

Hoege, C., Pfander, B., Moldovan, G.-L., Pyrowolakis, G., Jentsch, S., 2002. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature 419, 135–141.

Hoffmann, G.R., 1980. Genetic effects of dimethyl sulfate, diethyl sulfate, and related compounds. Mutation Research 75, 63–129.

Hsieh, P., Yamane, K., 2008. DNA mismatch repair: Molecular mechanism, cancer, and ageing. Mechanisms of ageing and development 129, 391–407.

Hu, J., Guo, L., Wu, K., Liu, B., Lang, S., Huang, L., 2012. Template-dependent polymerization across discontinuous templates by the heterodimeric primase from the hyperthermophilic archaeon Sulfolobus solfataricus. Nucleic Acids Res. 40, 3470–3483.

Huang, J.C., Svoboda, D.L., Reardon, J.T., Sancar, A., 1992. Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5“ and the 6th phosphodiester bond 3” to the photodimer. Proc. Natl. Acad. Sci. USA 89, 3664–3668.

Huberman, J.A., Riggs, A.D., 1966. Autoradiography of chromosomal DNA fibers from Chinese hamster cells. Proc. Natl. Acad. Sci. USA 55, 599–606.

Huberman, J.A., Riggs, A.D., 1968. On the mechanism of DNA replication in mammalian chromosomes. J. Mol. Biol. 32, 327–341.

Hübscher, U., Maga, G., 2011. DNA replication and repair bypass machines. Current Opinion in Chemical Biology 15, 627–635.

Ilves, I., Petojevic, T., Pesavento, J.J., Botchan, M.R., 2010. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol. Cell 37, 247–258.

Im, J.-S., Ki, S.-H., Farina, A., Jung, D.-S., Hurwitz, J., Lee, J.-K., 2009. Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. Proc. Natl. Acad. Sci. USA 106, 15628–15632.

Indiani, C., Patel, M., Goodman, M.F., O'Donnell, M.E., 2013. RecA acts as a switch to regulate polymerase occupancy in a moving replication fork. Proc. Natl. Acad. Sci. USA 110, 5410–5415.

Ito, N., Nureki, O., Shirouzu, M., Yokoyama, S., Hanaoka, F., 2003. Crystal structure of the Pyrococcus horikoshii DNA primase-UTP complex: implications for the mechanism of primer synthesis. Genes Cells 8, 913–923.

Iyer, L.M., Koonin, E.V., Leipe, D.D., Aravind, L., 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. 33, 3875–3896.

Iyer, R.R., Pluciennik, A., Burdett, V., Modrich, P.L., 2006. DNA mismatch repair: functions and mechanisms. Chem. Rev. 106, 302–323.

J Ito, D.K.B., 1991. Compilation and alignment of DNA polymerase sequences. Nucleic Acids Res. 19, 4045.

Jackson, D.A., Pombo, A., 1998. Replicon clusters are stable units of chromosome

structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. J. Cell Biol. 140, 1285–1295.
Jemt, E., Farge, G., Bäckström, S., Holmlund, T., Gustafsson, C.M., Falkenberg, M., 2011. The mitochondrial DNA helicase TWINKLE can assemble on a closed circular template and support initiation of DNA synthesis. Nucleic Acids Res. 39, 9238–9249.
Jiang, W., McDonald, D., Hope, T.J., Hunter, T., 1999a. Mammalian Cdc7-Dbf4 protein kinase complex is essential for initiation of DNA replication. The EMBO Journal 18, 5703–5713.
Jiang, W., Wells, N.J., Hunter, T., 1999b. Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6. Proc. Natl. Acad. Sci. USA 96, 6193–6198.
Johnson, R.E., Haracska, L., Prakash, S., Prakash, L., 2001. Role of DNA polymerase η in the bypass of a (6-4) TT photoproduct. Mol. Cell. Biol. 21, 3558–3563.
Johnson, S.J., Beese, L.S., 2004. Structures of mismatch replication errors observed in a DNA polymerase. Cell 116, 803–816.
Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202.
Jonsson, Z.O., Mossi, R., Allen, B.L., 1997. Replication Factor C Interacts with the C-terminal Side of Proliferating Cell Nuclear Antigen. J. Biol. Chem. 272, 1769–1776.
Joo, W., Xu, G., Persky, N.S., Smogorzewska, A., Rudge, D.G., Buzovetsky, O., Elledge, S.J., Pavletich, N.P., 2011. Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway. Science 333, 312–316.
Joyce, C.M., Steitz, T.A., 1994. Function and structure relationships in DNA polymerases. Annual Review of Biochemistry 63, 777–822.
Jozwiakowski, S.K., Borazjani Gholami, F., Doherty, A.J., 2015. Archaeal replicative primases can perform translesion DNA synthesis. Proc. Natl. Acad. Sci. USA.
Jozwiakowski, S.K., Connolly, B.A., 2011. A modified family-B archaeal DNA polymerase with reverse transcriptase activity. ChemBioChem 12, 35–37.
Kaina, B., Christmann, M., Naumann, S., Roos, W.P., 2007. MGMT: Key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. DNA repair 6, 1079–1099.
Kalifa, L., Sia, E.A., 2007. Analysis of Rev1p and Pol zeta in mitochondrial mutagenesis suggests an alternative pathway of damage tolerance. DNA repair 6, 1732–1739.
Kanno, S.-I., Kuzuoka, H., Sasao, S., Hong, Z., Lan, L., Nakajima, S., Yasui, A., 2007. A novel human AP endonuclease with conserved zinc-finger-like motifs involved in DNA strand break responses. The EMBO Journal 26, 2094–2103.
Kannouche, P.L., Wing, J., Lehmann, A.R., 2004. Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. Mol. Cell 14, 491–500.
Kasiviswanathan, R., Gustafson, M.A., Copeland, W.C., Meyer, J.N., 2012. Human mitochondrial DNA polymerase γ exhibits potential for bypass and mutagenesis at UV-induced cyclobutane thymine dimers. J. Biol. Chem. 287, 9222–9229.
Kazlauskas, D., Venclovas, C., 2014. Herpesviral helicase-primase subunit UL8 is inactivated B-family polymerase. Bioinformatics 30, 2093–2097.
Keen, B.A., Jozwiakowski, S.K., Bailey, L.J., Bianchi, J., Doherty, A.J., 2014. Molecular dissection of the domain architecture and catalytic activities of human PrimPol. Nucleic Acids Res. 42, 5830–5845.
Kelman, Z., White, M.F., 2005. Archaeal DNA replication and repair. Curr. Opin. Microbiol. 8, 669–676.
Kelner, A., 1949. Effect of Visible Light on the Recovery of Streptomyces Griseus Conidia from Ultra-violet Irradiation Injury. Proc. Natl. Acad. Sci. USA 35, 73–79.

Kennedy, R.D., D'Andrea, A.D., 2006. DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. J. Clin. Oncol. 24, 3799–3808.
Kenny, M.K., Lee, S.H., Hurwitz, J., 1989. Multiple functions of human single-stranded-DNA binding protein in simian virus 40 DNA replication: single-strand stabilization and stimulation of DNA polymerases alpha and delta. Proc. Natl. Acad. Sci. USA 86, 9757–9761.
Kent, T., Chandramouly, G., McDevitt, S.M., Ozdemir, A.Y., Pomerantz, R.T., 2015. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. Nature Structural & Molecular Biology.
Kim, J.-K., Choi, B.-S., 1995. The Solution Structure of DNA Duplex-Decamer Containing the (6-4) Photoproduct of Thymidylyl (3′→ 5′) Thymidine by NMR and Relaxation Matrix Refinement. European Journal of Biochemistry 228, 849–854.
Kim, N., Jinks-Robertson, S., 2009. dUTP incorporation into genomic DNA is linked to transcription in yeast. Nature 459, 1150–1153.
King, M., Attardi, G., 1989. Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation. Science 246, 500–503.
Kirk, B.W., Kuchta, R.D., 1999a. Human DNA primase: anion inhibition, manganese stimulation, and their effects on in vitro start-site selection. Biochemistry 38, 10126–10134.
Kirk, B.W., Kuchta, R.D., 1999b. Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: primase and the family X polymerases share significant sequence homology. Biochemistry 38, 7727–7736.
Klapstein, K., Chou, T., Bruinsma, R., 2004. Physics of RecA-mediated homologous recognition. Biophys. J. 87, 1466–1477.
Klenow, H., Overgaard-Hansen, K., 1970. Proteolytic cleavage of DNA polymerase from Escherichia Coli B into an exonuclease unit and a polymerase unit. FEBS Lett. 6, 25–27.
Klinedinst, D.K., Challberg, M.D., 1994. Helicase-primase complex of herpes simplex virus type 1: a mutation in the UL52 subunit abolishes primase activity. J. Virol. 68, 3693–3701.
Klinge, S., Hirst, J., Maman, J.D., Krude, T., Pellegrini, L., 2007. An iron-sulfur domain of the eukaryotic primase is essential for RNA primer synthesis. Nature Structural & Molecular Biology 14, 875–877.
Koonin, E.V., Wolf, Y.I., Kondrashov, A.S., Aravind, L., 2000. Bacterial homologs of the small subunit of eukaryotic DNA primase. J. Mol. Microbiol. Biotechnol. 2, 509–512.
Korhonen, J.A., Gaspari, M., Falkenberg, M., 2003. TWINKLE Has 5′ → 3′ DNA Helicase Activity and Is Specifically Stimulated by Mitochondrial Single-stranded DNA-binding Protein. J. Biol. Chem. 278, 48627–48632.
Korhonen, J.A., Pham, X.H., Pellegrini, M., Falkenberg, M., 2004. Reconstitution of a minimal mtDNA replisome in vitro. The EMBO Journal 23, 2423–2429.
Kornberg, A., 1962. Horizons in biochemistry. Academic Press, New York.
Kuchta, R.D., Stengel, G., 2010. Mechanism and evolution of DNA primases. Biochim. Biophys. Acta 1804, 1180–1189.
Kukimoto, I., Igaki, H., Kanda, T., 2001. Human CDC45 protein binds to minichromosome maintenance 7 protein and the p70 subunit of DNA polymerase α. European Journal of Biochemistry 265, 936–943.
Kunkel, T.A., 2011. Balancing eukaryotic replication asymmetry with replication fidelity. Current Opinion in Chemical Biology 15, 620–626.
Kunkel, T.A., Burgers, P.M., 2008. Dividing the workload at a eukaryotic replication fork. Trends Cell Biol. 18, 521–527.
Kuraoka, I., Kobertz, W.R., Ariza, R.R., Biggerstaff, M., Essigmann, J.M., Wood, R.D.,

2000. Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. J. Biol. Chem. 275, 26632–26636.
Kusumoto, R., Dawut, L., Marchetti, C., Lee, J.W., 2008. Biochemistry - Werner Protein Cooperates with the XRCC4-DNA Ligase IV Complex in End-Processing. Biochemistry 47, 7548–7556.
Kusumoto, R., Masutani, C., Iwai, S., Hanaoka, F., 2002. Translesion synthesis by human DNA polymerase eta across thymine glycol lesions. Biochemistry 41, 6090–6099.
Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.
Lao-Sirieix, S.-H., Bell, S.D., 2004. The heterodimeric primase of the hyperthermophilic archaeon Sulfolobus solfataricus possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. J. Mol. Biol. 344, 1251–1263.
Lao-Sirieix, S.-H., Nookala, R.K., Roversi, P., Bell, S.D., Pellegrini, L., 2005. Structure of the heterodimeric core primase. Nature Structural & Molecular Biology 12, 1137–1144.
Laskowski, R.A., Swindells, M.B., 2011. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51, 2778–2786.
Le Breton, M., Henneke, G., Norais, C., Flament, D., Myllykallio, H., Querellou, J., Raffin, J.P., 2007. The heterodimeric primase from the euryarchaeon Pyrococcus abyssi: a multifunctional enzyme for initiation and repair? J. Mol. Biol. 374, 1172–1185.
Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J., Bensimon, A., 2006. DNA replication origin interference increases the spacing between initiation events in human cells. Mol. Biol. Cell 17, 5337–5345.
Lee, J.K., 2003. The Cdc23 (Mcm10) protein is required for the phosphorylation of minichromosome maintenance complex by the Dfp1-Hsk1 kinase. Proc. Natl. Acad. Sci. USA 100, 2334–2339.
Lehman, I.R., 1974. DNA ligase: structure, mechanism, and function. Science 186, 790–797.
Lehman, I.R., Bessman, M.J., Simms, E.S., Kornberg, A., 1958. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. J. Biol. Chem. 233, 163–170.
Lehmann, A.R., 2003. DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. Biochimie 85, 1101–1111.
Lehmann, A.R., Niimi, A., Ogi, T., Brown, S., Sabbioneda, S., Wing, J.F., Kannouche, P.L., Green, C.M., 2007. Translesion synthesis: Y-family polymerases and the polymerase switch. DNA repair 6, 891–899.
Leipe, D.D., Aravind, L., Koonin, E.V., 1999. Did DNA replication evolve twice independently? Nucleic Acids Res. 27, 3389–3401.
Levy, R., Weiss, R., Chen, G., Iverson, B.L., Georgiou, G., 2001. Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones. Protein Expression and Purification 23, 338–347.
Li, X., Burgers, P.M., 1994. Molecular cloning and expression of the Saccharomyces cerevisiae RFC3 gene, an essential component of replication factor C. Proc. Natl. Acad. Sci. USA 91, 868–872.
Li, X., Li, J., Harrington, J., Lieber, M.R., Burgers, P.M.J., 1995. Lagging Strand DNA Synthesis at the Eukaryotic Replication Fork Involves Binding and Stimulation of FEN-1 by Proliferating Cell Nuclear Antigen. J. Biol. Chem. 270, 22109–22112.
Li, Y., Kong, Y., Korolev, S., Waksman, G., 1998a. Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with

deoxyribonucleoside triphosphates. Protein Sci. 7, 1116–1123.
Li, Y., Korolev, S., Waksman, G., 1998b. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. The EMBO Journal 17, 7514–7525.
Li, Y., Waksman, G., 2001. Crystal structures of a ddATP-, ddTTP-, ddCTP, and ddGTP- trapped ternary complex of Klentaq1: Insights into nucleotide incorporation and selectivity. Protein Sci. 10, 1225–1233.
Liang, Q., Dexheimer, T.S., Zhang, P., Rosenthal, A.S., Villamil, M.A., You, C., Zhang, Q., Chen, J., Ott, C.A., Sun, H., Luci, D.K., Yuan, B., Simeonov, A., Jadhav, A., Xiao, H., Wang, Y., Maloney, D.J., Zhuang, Z., 2014. A selective USP1–UAF1 inhibitor links deubiquitination to DNA damage responses. Nat Chem Biol 10, 298–304.
Lim, S.E., 1999. The Mitochondrial p55 Accessory Subunit of Human DNA Polymerase gamma Enhances DNA Binding, Promotes Processive DNA Synthesis, and Confers N-Ethylmaleimide Resistance. J. Biol. Chem. 274, 38197–38203.
Lindahl, T., Nyberg, B., 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry 13, 3405–3410.
Ling, H., Boudsocq, F., Woodgate, R., Yang, W., 2001. Crystal Structure of a Y-Family DNA Polymerase in Action: A Mechanism for Error-Prone and Lesion-Bypass Replication. Cell 107, 91–102.
Lipps, G., Röther, S., Hart, C., Krauss, G., 2003. A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. The EMBO Journal 22, 2516–2525.
Lipps, G., Weinzierl, A.O., Scheven, von, G., Buchen, C., Cramer, P., 2004. Structure of a bifunctional DNA primase-polymerase. Nature Structural & Molecular Biology 11, 157–162.
Liu, L., Huang, M., 2015. Essential role of the iron-sulfur cluster binding domain of the primase regulatory subunit Pri2 in DNA replication initiation. Protein Cell.
Liu, L., Komori, K., Sonoko, I., Bocquier, A.A., Cann, I.K.O., Kohda, D., Ishino, Y., 2001. The Archaeal DNA Primase. Biochemical Characterization of the P41-P46 Complex From Pyroccocus Furiosis. J. Biol. Chem. 276, 45484–45490.
Liu, S., Ho, C.K., Ouyang, J., Zou, L., 2013. Nek1 kinase associates with ATR-ATRIP and primes ATR for efficient DNA damage signaling. Proc. Natl. Acad. Sci. USA 110, 2175–2180.
Liu, Z., Ding, Y., Du, A., Zhang, B., Zhao, G., Meiping, D., 2008. A novel Twinkle (PEO1) gene mutation in a Chinese family with adPEO. Molecular Vision 14, 1995.
Lobley, A., Whitmore, L., Wallace, B.A., 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. Bioinformatics 18, 211–212.
Loeb, L.A., 1991. Mutator phenotype may be required for multistage carcinogenesis. Cancer Res. 51, 3075–3079.
Loeb, L.A., Preston, B.D., 1986. Mutagenesis by apurinic/apyrimidinic sites. Annual review of genetics 20, 201–230.
Lukianova, O.A., David, S.S., 2005. A role for iron–sulfur clusters in DNA repair. Current Opinion in Chemical Biology 9, 145–151.
Lutzmann, M., Maiorano, D., Méchali, M., 2006. A Cdt1–geminin complex licenses chromatin for DNA replication and prevents rereplication during S phase in Xenopus. The EMBO Journal 25, 5764–5774.
Lutzmann, M., Méchali, M., 2008. MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. Mol. Cell 31, 190–200.
Machwe, A., Xiao, L., Groden, J., Orren, D.K., 2006. The Werner and Bloom syndrome

proteins catalyze regression of a model replication fork. Biochemistry 45, 13939–13946.

Maga, G., Villani, G., Crespan, E., Wimmer, U., Ferrari, E., Bertocci, B., Hübscher, U., 2007. 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. Nature 447, 606–608.

Maga, G., Villani, G., Tillement, V., Stucki, M., Locatelli, G.A., Frouin, I., Spadari, S., Hübscher, U., 2001. Okazaki fragment processing: Modulation of the strand displacement activity of DNA polymerase by the concerted action of replication protein A, proliferating cell nuclear antigen, and flap endonuclease-1. Proc. Natl. Acad. Sci. USA 98, 14298–14303.

Mahmood, N., Khan, S.G., Athar, M., Rahman, Q., 1994. Differential Role of Hydrogen Peroxide and Organic Peroxides Augmenting Asbestos-Mediated DNA Damage: Implications for Asbestos Induced Carcinogenesis. Biochem. Biophys. Res. Comm. 200, 687–694.

Majka, J., Binz, S.K., Wold, M.S., Burgers, P.M.J., 2006. Replication protein A directs loading of the DNA damage checkpoint clamp to 5'-DNA junctions. J. Biol. Chem. 281, 27855–27861.

Maki, H., Sekiguchi, M., 1992. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. Nature 355, 273–275.

Malu, S., Malshetty, V., Francis, D., Cortes, P., 2012. Role of non-homologous end joining in V(D)J recombination. Immunologic Research 54, 233–246.

Mari, P.O., Florea, B.I., Persengiev, S.P., Verkaik, N.S., Bruggenwirth, H.T., Modesti, M., Giglia-Mari, G., Bezstarosti, K., Demmers, J.A.A., Luider, T.M., Houtsmuller, A.B., van Gent, D.C., 2006. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. Proc. Natl. Acad. Sci. USA 103, 18597–18602.

Maric, M., Maculins, T., De Piccoli, G., Labib, K., 2014. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. Science 346, 1253596.

Marsden, H.S., McLean, G.W., Barnard, E.C., Francis, G.J., MacEachran, K., Murphy, M., McVey, G., Cross, A., Abbotts, A.P., Stow, N.D., 1997. The catalytic subunit of the DNA polymerase of herpes simplex virus type 1 interacts specifically with the C terminus of the UL8 component of the viral helicase-primase complex. J. Virol. 71, 6390–6397.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., Hanaoka, F., 1999. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. Nature 399, 700–704.

Matsuda, T., Bebenek, K., Masutani, C., Rogozin, I.B., Hanaoka, F., Kunkel, T.A., 2001. Error rate and specificity of human and murine DNA polymerase eta. J. Mol. Biol. 312, 335–346.

Matsui, E., Nishio, M., Yokoyama, H., Harata, K., Darnis, S., Matsui, I., 2003. Distinct domain functions regulating de novo DNA synthesis of thermostable DNA primase from hyperthermophile Pyrococcus horikoshii. Biochemistry 42, 14968–14976.

Matsunaga, T., Park, C.H., Bessho, T., Mu, D., Sancar, A., 1996. Replication Protein A Confers Structure-specific Endonuclease Activities to the XPF-ERCC1 and XPG Subunits of Human DNA Repair Excision Nuclease. J. Biol. Chem. 271, 11047–11050.

McNicholas, S., Potterton, E., Wilson, K.S., Noble, M.E.M., 2011. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr. D Biol. Crystallogr. 67, 386–394.

Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C.J., Edwards, A.M., Chazin, W.J., 2000. Structural basis for the recognition of DNA repair

proteins UNG2, XPA, and RAD52 by replication factor RPA. Cell 103, 449–456.
Meselson, M., Stahl, F.W., 1958. The Replication of Dna in Escherichia Coli. Proc. Natl. Acad. Sci. USA 44, 671–682.
Mikhailov, V.S., Rohrmann, G.F., 2002. Baculovirus replication factor LEF-1 is a DNA primase. J. Virol. 76, 2287–2297.
Mimitou, E.P., Symington, L.S., 2009. DNA end resection: many nucleases make light work. DNA repair 8, 983–995.
Mimori, T., Hardin, J.A., Steitz, J.A., 1986. Characterization of the DNA-binding protein antigen Ku recognized by autoantibodies from patients with rheumatic disorders. J. Biol. Chem. 261, 2274–2278.
Mimura, S., Takisawa, H., 1998. Xenopus Cdc45-dependent loading of DNA polymerase alpha onto chromatin under the control of S-phase cdk. The EMBO Journal 17, 5699–5707.
Minato, S., Werbin, H., 1972. Excitation and fluorescence spectra of the chromophore associated with the DNA-photoreactivating enzyme from the blue-green alga Anacystis nidulans. Photochem. Photobiol. 15, 97–100.
Minotti, G., Menna, P., Salvatorelli, E., Cairo, G., Gianni, L., 2004. Anthracyclines: Molecular Advances and Pharmacologic Developments in Antitumor Activity and Cardiotoxicity. Pharmacological Reviews 56, 185–229.
Miyabe, I., Kunkel, T.A., Carr, A.M., 2011. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. PLoS genetics 7, e1002407.
Monnat, R.J., 2010. Human RECQ helicases: roles in DNA metabolism, mutagenesis and cancer biology. Semin. Cancer Biol. 20, 329–339.
Morgan, I.G., Ohno-Matsui, K., Saw, S.M., 2012. Myopia. The Lancet 379, 1739–1748.
Morrison, A., Araki, H., Clark, A.B., Hamatake, R.K., Sugino, A., 1990. A third essential DNA polymerase in S. cerevisiae. Cell 62, 1143–1151.
Morrison, A.J., Shen, X., 2009. Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. Nature Reviews Molecular Cell Biology 10, 373–384.
Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L.H.F., Fousteri, M.I., 2007. Sealing of Chromosomal DNA Nicks during Nucleotide Excision Repair Requires XRCC1 and DNA Ligase IIIα in a Cell-Cycle-Specific Manner. Mol. Cell 27, 311–323.
Mourón, S., Rodriguez-Acebes, S., Martínez-Jiménez, M.I., Garcia-Gomez, S., Chocron, S., Blanco, L., Mendez, J., 2013. Repriming of DNA synthesis at stalled replication forks by human PrimPol. Nature Structural & Molecular Biology 20, 1383–1389.
Muller, H.J., 1927. Artificial Transmutation of the Gene. Science 66, 84–87.
Myers, T.W., Gelfand, D.H., 1991. Reverse transcription and DNA amplification by a Thermus thermophilus DNA polymerase. Biochemistry 30, 7661–7666.
Nair, P.A., Smith, P., Shuman, S., 2010. Structure of bacterial LigD 3′-phosphoesterase unveils a DNA repair superfamily. Proc. Natl. Acad. Sci. USA 107, 12822–12827.
Nakamura, J., Walker, V.E., Upton, P.B., Chiang, S.-Y., Kow, Y.W., Swenberg, J.A., 1998. Highly Sensitive Apurinic/Apyrimidinic Site Assay Can Detect Spontaneous and Chemically Induced Depurination under Physiological Conditions. Cancer Res. 58, 222–225.
Nemec, A.A., Donigan, K.A., Murphy, D.L., Jaeger, J., Sweasy, J.B., 2012. Colon cancer-associated DNA polymerase β variant induces genomic instability and cellular transformation. J. Biol. Chem. 287, 23840–23849.
Nemec, A.A., Murphy, D.L., Donigan, K.A., Sweasy, J.B., 2014. The S229L Colon Tumor-Associated Variant of DNA Polymerase Beta Induces Cellular

Transformation As A Result Of Decreased Polymerization Efficiency. J. Biol. Chem. 289, 13708–13716.
New, J.H., Sugiyama, T., Zaitseva, E., Kowalczykowski, S.C., 1998. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. Nature 391, 407–410.
Newlon, C.S., Theis, J.F., 1993. The structure and function of yeast ARS elements. Curr. Opin. Genet. Dev. 3, 752–758.
Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M.J., Kunkel, T.A., 2008. Division of labor at the eukaryotic replication fork. Mol. Cell 30, 137–144.
Nick McElhinny, S.A., Havener, J.M., García-Díaz, M., Juarez, R., Bebenek, K., Kee, B.L., Blanco, L., Kunkel, T.A., Ramsden, D.A., 2005. A Gradient of Template Dependence Defines Distinct Biological Roles for Family X Polymerases in Nonhomologous End Joining. Mol. Cell 19, 357–366.
Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., O'Neal, C., 1965. RNA codewords and protein synthesis, VII. On the general nature of the RNA code. Proc. Natl. Acad. Sci. USA 53, 1161–1168.
O'Donovan, A., Davies, A.A., Moggs, J.G., West, S.C., Wood, R.D., 1994. XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair. Nature 371, 432–435.
Ohashi, E., Bebenek, K., Matsuda, T., Feaver, W.J., Gerlach, V.L., Friedberg, E.C., Ohmori, H., Kunkel, T.A., 2000. Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene. J. Biol. Chem. 275, 39678–39684.
Olive, P.L., 1998. The Role of DNA Single- and Double-Strand Breaks in Cell Killing by Ionizing Radiation. Radiation research 150, S42–S51.
Oliveira, M.T., Kaguni, L.S., 2010. Functional roles of the N- and C-terminal regions of the human mitochondrial single-stranded DNA-binding protein. PloS one 5, e15379.
Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., Steitz, T.A., 1985. Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. Nature 313, 762–766.
Ong, J.L., Loakes, D., Jaroslawski, S., Too, K., Holliger, P., 2006. Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. J. Mol. Biol. 361, 537–550.
Park, H., Zhang, K., Ren, Y., Nadji, S., Sinha, N., Taylor, J.-S., Kang, C., 2002. Crystal structure of a DNA decamer containing a cis-syn thymine dimer. Proc. Natl. Acad. Sci. USA 99, 15965–15970.
Park, H.W., Kim, S.T., Sancar, A., Deisenhofer, J., 1995. Crystal structure of DNA photolyase from Escherichia coli. Science 268, 1866–1872.
Pascal, J.M., O'Brien, P.J., Tomkinson, A.E., Ellenberger, T., 2004. Human DNA ligase I completely encircles and partially unwinds nicked DNA. Nature 432, 473–478.
Pata, J.D., 2010. Structural diversity of the Y-family DNA polymerases. Biochim. Biophys. Acta 1804, 1124–1135.
Patel, P.K., Arcangioli, B., Baker, S.P., Bensimon, A., Rhind, N., 2005. DNA Replication Origins Fire Stochastically in Fission Yeast. Mol. Biol. Cell 17, 308–316.
Patil, M., Pabla, N., Ding, H.-F., Dong, Z., 2013. Nek1 interacts with Ku80 to assist chromatin loading of replication factors and S-phase progression. Cell Cycle 12, 2608–2616.
Pavlov, Y.I., Frahm, C., Nick McElhinny, S.A., Niimi, A., Suzuki, M., Kunkel, T.A., 2006. Evidence that Errors Made by DNA Polymerase α are Corrected by DNA Polymerase δ. Current Biology 16, 202–207.

Pavlov, Y.I., Shcherbakova, P.V., 2010. DNA polymerases at the eukaryotic fork-20 years later. Mutation Research 685, 45–53.

Pelletier, H., Sawaya, M.R., Wolfle, W., Wilson, S.H., Kraut, J., 1996. A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase beta. Biochemistry 35, 12762–12777.

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567.

Perry, J.J.P., Yannone, S.M., Holden, L.G., Hitomi, C., Asaithamby, A., Han, S., Cooper, P.K., Chen, D.J., Tainer, J.A., 2006. WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. Nature Structural & Molecular Biology 13, 414–422.

Petermann, E., Helleday, T., 2010. Pathways of mammalian replication fork restart. Nature Reviews Molecular Cell Biology 11, 683–687.

Petermann, E., Orta, M.L., Issaeva, N., Schultz, N., Helleday, T., 2010. Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. Mol. Cell 37, 492–502.

Petruska, J., Goodman, M.F., Boosalis, M.S., Sowers, L.C., Cheong, C., Tinoco, I., 1988. Comparison between DNA melting thermodynamics and DNA polymerase fidelity. Proc. Natl. Acad. Sci. USA 85, 6252–6256.

Pfeifer, G.P., 2008. Formation and Processing of UV Photoproducts: Effects of DNA Sequence and Chromatin Environment. Photochem. Photobiol. 65, 270–283.

Pham, X.H., Farge, G., Shi, Y., Gaspari, M., Gustafsson, C.M., Falkenberg, M., 2006. Conserved sequence box II directs transcription termination and primer formation in mitochondria. J. Biol. Chem. 281, 24647–24652.

Pinz, K.G., Shibutani, S., Bogenhagen, D.F., 1995. Action of mitochondrial DNA polymerase gamma at sites of base loss or oxidative damage. J. Biol. Chem. 270, 9202–9206.

Pitcher, R.S., Brissett, N.C., Picher, A.J., Andrade, P., Juarez, R., Thompson, D., Fox, G.C., Blanco, L., Doherty, A.J., 2007. Structure and function of a mycobacterial NHEJ DNA repair polymerase. J. Mol. Biol. 366, 391–405.

Pitcher, R.S., Tonkin, L.M., Green, A.J., Doherty, A.J., 2005. Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. J. Mol. Biol. 351, 531–544.

Poot, R.A., Dellaire, G., Hülsmann, B.B., Grimaldi, M.A., Corona, D.F., Becker, P.B., Bickmore, W.A., Varga-Weisz, P.D., 2000. HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. The EMBO Journal 19, 3377–3387.

Popanda, O., Thielmann, H.W., 1992. The function of DNA polymerases in DNA repair synthesis of ultraviolet-irradiated human fibroblasts. Biochim. Biophys. Acta 1129, 155–160.

Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., Noble, M., 2002. The CCP4 molecular-graphics project. Acta Crystallogr. D Biol. Crystallogr. 58, 1955–1957.

Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G.N., Cohen, S., Perrakis, A., Noble, M., 2004. Developments in the CCP4 molecular-graphics project. Acta Crystallogr. D Biol. Crystallogr. 60, 2288–2294.

Povirk, L.F., Zhou, T., Zhou, R., Cowan, M.J., Yannone, S.M., 2007. Processing of 3′-Phosphoglycolate-terminated DNA Double Strand Breaks by Artemis Nuclease. J. Biol. Chem. 282, 3547–3558.

Prakash, L., Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S., 2000. Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. Nature

406, 1015–1019.
Prasad, R., 2001. DNA Polymerase beta -mediated Long Patch Base Excision Repair. Poly(ADP-Ribose) Polymerase-1 Stimulates Displacement DNA Synthesis. J. Biol. Chem. 276, 32411–32414.
Prasad, R., Dianov, G.L., Bohr, V.A., Wilson, S.H., 2000. FEN1 Stimulation of DNA Polymerase β Mediates an Excision Step in Mammalian Long Patch Base Excision Repair. J. Biol. Chem. 275, 4460–4466.
Prat, F., Houk, K.N., Foote, C.S., 1998. Effect of Guanine Stacking on the Oxidation of 8-Oxoguanine in B-DNA. J. Am. Chem. Soc. 120, 845–846.
Prato, S., Vitale, R.M., Contursi, P., Lipps, G., Saviano, M., Rossi, M., Bartolucci, S., 2008. Molecular modeling and functional characterization of the monomeric primase-polymerase domain from the Sulfolobus solfataricus plasmid pIT3. FEBS J. 275, 4389–4402.
Prelich, G., Kostura, M., Marshak, D.R., Mathews, M.B., Stillman, B., 1987. The cell-cycle regulated proliferating cell nuclear antigen is required for SV40 DNA replication in vitro. Nature 326, 471–475.
Pursell, Z.F., Isoz, I., Lundström, E.-B., Johansson, E., Kunkel, T.A., 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science 317, 127–130.
Rao, P.N., Johnson, R.T., 1970. Mammalian Cell Fusion : Studies on the Regulation of DNA Synthesis and Mitosis. Nature 225, 159–164.
Rastogi, R.P., Richa, Kumar, A., Tyagi, M.B., Sinha, R.P., 2010. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. J Nucleic Acids 592980, 1–31.
Reardon, J.T., Vaisman, A., Chaney, S.G., Sancar, A., 1999. Efficient Nucleotide Excision Repair of Cisplatin, Oxaliplatin, and Bis-aceto-ammine-dichloro-cyclohexylamine-platinum(IV) (JM216) PlatinumIntrastrand DNA Diadducts. Cancer Res. 59, 3968–3971.
Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., Diffley, J.F.X., 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell 139, 719–730.
Ricke, R.M., Bielinsky, A.-K., 2004. Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. Mol. Cell 16, 173–185.
Rink, S.M., Solomon, M.S., Taylor, M.J., Rajur, S.B., McLaughlin, L.W., Hopkins, P.B., 1993. Covalent structure of a nitrogen mustard-induced DNA interstrand cross-link: an N7-to-N7 linkage of deoxyguanosine residues at the duplex sequence 5'-d(GNC). J. Am. Chem. Soc. 115, 2551–2557.
Ropp, P.A., Copeland, W.C., 1996. Cloning and Characterization of the Human Mitochondrial DNA Polymerase, DNA Polymerase γ. Genomics 36, 449–458.
Rose, G., Romeo, G., Dato, S., Crocco, P., Bruni, A.C., Hervonen, A., Majamaa, K., Sevini, F., Franceschi, C., Passarino, G., GEnetics of Healthy Ageing Project Consortium, 2010. Somatic point mutations in mtDNA control region are influenced by genetic background and associated with healthy aging: a GEHA study. PloS one 5, e13395.
Rosenberg, B., Vancamp, L., Trosko, J.E., Mansour, V.H., 1969. Platinum Compounds: a New Class of Potent Antitumour Agents. Nature 222, 385–386.
Rothwell, P.J., Mitaksov, V., Waksman, G., 2005. Motions of the Fingers Subdomain of Klentaq1 Are Fast and Not Rate Limiting: Implications for the Molecular Basis of Fidelity in DNA Polymerases. Mol. Cell 19, 345–355.
Rudd, S.G., 2013. Cellular and biochemical characterisation of PrimPol, a novel eukaryotic primase-polymerase involved in DNA damage tolerance. University of Sussex.

Rudd, S.G., Bianchi, J., Doherty, A.J., 2014. PrimPol–a new polymerase on the block. Molecular & Cellular Oncology 1, e960754.

Rudd, S.G., Glover, L., Jozwiakowski, S.K., Horn, D., Doherty, A.J., 2013. PPL2 translesion polymerase is essential for the completion of chromosomal DNA replication in the African Trypanosome. Mol. Cell 52, 554–565.

Sambrook, J., Russell, D.W., 2006. SDS-Polyacrylamide Gel Electrophoresis of Proteins. Cold Spring Harbor Protocols doi:10.1101–pdb.prot4540.

Samuels, M., Gulati, G., Shin, J.-H., Opara, R., McSweeney, E., Sekedat, M., Long, S., Kelman, Z., Jeruzalmi, D., 2009. A biochemically active MCM-like helicase in Bacillus cereus. Nucleic Acids Res. 37, 4441–4452.

Sancar, A., 2003. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. Chem. Rev. 103, 2203–2237.

Sancar, G.B., 1990. DNA photolyases: physical properties, action mechanism, and roles in dark repair. Mutation Research 236, 147–160.

Sanchez-Berrondo, J., Mesa, P., Ibarra, A., Martínez-Jiménez, M.I., Blanco, L., Mendez, J., Boskovic, J., Montoya, G., 2012. Molecular architecture of a multifunctional MCM complex. Nucleic Acids Res. 40, 1366–1380.

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D., Smith, S., Uziel, T., Sfez, S., et, A., 1995. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 268, 1749–1753.

Saw, S.-M., Gazzard, G., Shih-Yen, E.C., Chua, W.-H., 2005. Myopia and associated pathological complications. Oph Phys Optics 25, 381–391.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Mizrachi, I., Ostell, J., Panchenko, A., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y., John Wilbur, W., Yaschenko, E., Ye, J., 2010. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 38, D5–16.

Schmitt, M.W., Matsumoto, Y., Loeb, L.A., 2009. High fidelity and lesion bypass capability of human DNA polymerase delta. Biochimie 91, 1163–1172.

Schofield, M.J., Brownewell, F.E., Nayak, S., Du, C., Kool, E.T., Hsieh, P., 2001. The Phe-X-Glu DNA Binding Motif of MutS. The role of Hydrogen Bonding in Mismatch Recognition. J. Biol. Chem. 276, 45505–45508.

Seroussi, E., Lavi, S., 1993. Replication protein A is the major single-stranded DNA binding protein detected in mammalian cell extracts by gel retardation assays and UV cross-linking of long and short single-stranded DNA molecules. J. Biol. Chem. 268, 7147–7154.

Shannon, R.D., 1976. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. Acta Crystallographica A 32, 751–767.

Shapiro, T.A., Englund, P.T., 1995. The structure and replication of kinetoplast DNA. Annual Reviews in Microbiology 49, 117–143.

Shcherbakova, P.V., Pavlov, Y.I., 1996. 3“-->5” exonucleases of DNA polymerases epsilon and delta correct base analog induced DNA replication errors on opposite DNA strands in Saccharomyces cerevisiae. Genetics 142, 717–726.

Sherman, G., Gottlieb, J., Challberg, M.D., 1992. The UL8 subunit of the herpes simplex virus helicase-primase complex is required for efficient primer utilization. J. Virol. 66, 4884–4892.

Shivji, M., Kenny, M.K., Wood, R.D., 1992. Proliferating cell nuclear antigen is required for DNA excision repair. Cell 69, 367–374.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R.,

McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.

Simon, A.C., Zhou, J.C., Perera, R.L., van Deursen, F., Evrin, C., Ivanova, M.E., Kilkenny, M.L., Renault, L., Kjaer, S., Matak-Vinković, D., Labib, K., Costa, A., Pellegrini, L., 2014. A Ctf4 trimer couples the CMG helicase to DNA polymerase α in the eukaryotic replisome. Nature 510, 293–297.

Simon, M., Giot, L., Faye, G., 1991. The 3“ to 5” exonuclease activity located in the DNA polymerase delta subunit of Saccharomyces cerevisiae is required for accurate replication. The EMBO Journal 10, 2165.

Sims, P., Grover, P.L., Swaisland, A., Pal, K., Hewer, A., 1974. Metabolic activation of benzo(a)pyrene proceeds by a diol-epoxide. Nature 252, 326–328.

Simsek, D., Furda, A., Gao, Y., Artus, J., Brunet, E., Hadjantonakis, A.-K., Van Houten, B., Shuman, S., McKinnon, P.J., Jasin, M., 2011. Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. Nature 471, 245–248.

Soler, N., Marguet, E., Cortez, D., Desnoues, N., Keller, J., van Tilbeurgh, H., Sezonov, G., Forterre, P., 2010. Two novel families of plasmids from hyperthermophilic archaea encoding new families of replication proteins. Nucleic Acids Res. 38, 5088–5104.

Spelbrink, J.N., Li, F.-Y., Tiranti, V., Nikali, K., Yuan, Q.-P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G.-M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H.T., Zeviani, M., Larsson, C., 2001. Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. Nat. Genet. 28, 223–231.

Steitz, T.A., 1999. DNA polymerases: structural diversity and common mechanisms. J. Biol. Chem. 274, 17395–17398.

Steitz, T.A., Smerdon, S.J., Jäger, J., Joyce, C.M., 1994. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. Science 266, 2022–2025.

Stillman, B., 2008. DNA polymerases at the replication fork in eukaryotes. Mol. Cell 30, 259–260.

Stucki, M., Pascucci, B., Parlanti, E., Fortini, P., Wilson, S.H., 1998. Mammalian base excision repair by DNA polymerases delta and epsilon. Oncogene 17, 835–843.

Stura, E.A., Nemerow, G.R., Wilson, I.A., 1992. Strategies in the crystallization of glycoproteins and protein complexes. Journal of Crystal Growth 122, 273–285.

Sugasawa, K., Ng, J.M.Y., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P.M., Hanaoka, F., Bootsma, D., Hoeijmakers, J.H.J., 1998. Xeroderma Pigmentosum Group C Protein Complex Is the Initiator of Global Genome Nucleotide Excision Repair. Mol. Cell 2, 223–232.

Sundin, O., Varshavsky, A., 1981. Arrest of segregation leads to accumulation of highly intertwined catenated dimers: Dissection of the final stages of SV40 DNA replication. Cell 25, 659–669.

Svejstrup, J.Q., 2002. Mechanisms of transcription-coupled DNA repair. Nature Reviews Molecular Cell Biology 3, 21–29.

Sweetser, D., Nonet, M., Young, R.A., 1987. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc. Natl. Acad. Sci. USA 84, 1192–1196.

Tabor, S., Richardson, C.C., 1989. Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. Proc. Natl. Acad. Sci. USA 86, 4076–4080.

Takechi, S., Itoh, T., 1995. Initiation of unidirectional ColE2 DNA replication by a unique

priming mechanism. Nucleic Acids Res. 23, 4196–4201.
Takechi, S., Matsui, H., Itoh, T., 1995. Primer RNA synthesis by plasmid-specified Rep protein for initiation of ColE2 DNA replication. The EMBO Journal 14, 5141–5147.
Taneja, P., Nasheuer, H.-P., Hartmann, H., Grosse, F., Fanning, E., Weisshart, K., 2007. Timed interactions between viral and cellular replication factors during the initiation of SV40 in vitro DNA replication. Biochem. J. 407, 313–320.
Tang, J., Chu, G., 2002. Xeroderma pigmentosum complementation group E and UV-damaged DNA-binding protein. DNA repair 1, 601–616.
Taylor, J.-S., 1990. DNA, sunlight, and skin cancer. J. Chem. Educ. 67, 835–841.
Taylor, J.-S., Cohrs, M.P., 1987. DNA, light, and Dewar pyrimidinones: the structure and biological significance to TpT3. J. Am. Chem. Soc. 109, 2834–2835.
Todo, T., Ryo, H., Yamamoto, K., Toh, H., Inui, T., Ayaki, H., Nomura, T., Ikenaga, M., 1996. Similarity among the Drosophila (6-4)photolyase, a human photolyase homolog, and the DNA photolyase-blue-light photoreceptor family. Science 272, 109–112.
Topper, J.N., Clayton, D.A., 1990. Characterization of human MRP/Th RNA and its nuclear gene: full length MRP/Th RNA is an active endoribonuclease when assembled as an RNP. Nucleic Acids Res. 18, 793–799.
Tran, H.T., Gordenin, D.A., Resnick, M.A., 1999. The 3′→5′ Exonucleases of DNA Polymerases δ and ε and the 5′→3′ Exonuclease Exo1 Have Major Roles in Postreplication Mutation Avoidance in Saccharomyces cerevisiae. Mol. Cell. Biol. 19, 2000–2007.
Tsugita, A., Fraenkel-Conrat, H., 1960. The amino acid composition and C-terminal sequence of a chemically evoked mutant of TMV. Proc. Natl. Acad. Sci. USA 46, 636–642.
Tsurimoto, T., Stillman, B., 1989. Multiple replication factors augment DNA synthesis by the two eukaryotic DNA polymerases, alpha and delta. The EMBO Journal 8, 3883–3889.
Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., Pommier, Y., Tazi, J., Coquelle, A., Pasero, P., 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat. Cell Biol. 11, 1315–1324.
Tyynismaa, H., Sembongi, H., Bokori-Brown, M., Granycome, C., Ashley, N., Poulton, J., Jalanko, A., Spelbrink, J.N., Holt, I.J., Suomalainen, A., 2004. Twinkle helicase is essential for mtDNA maintenance and regulates mtDNA copy number. Hum. Mol. Genet. 13, 3219–3227.
Uchiyama, M., Griffiths, D., Arai, K.-I., Masia, H., 2001. Essential Role of Sna41/Cdc45 in Loading of DNA Polymerase alpha onto Minichromosome Maintenance Proteins in Fission Yeast. J. Biol. Chem. 276, 26189–26196.
Uematsu, N., Weterings, E., Yano, K.-I., Morotomi-Yano, K., Jakob, B., Taucher-Scholz, G., Mari, P.-O., van Gent, D.C., Benjamin P.C. Chen, Chen, D.J., 2007. Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. J. Cell Biol. 177, 219–229.
Uljon, S.N., Johnson, R.E., Edwards, T.A., Prakash, S., Prakash, L., Aggarwal, A.K., 2004. Crystal structure of the catalytic core of human DNA polymerase kappa. Structure 12, 1395–1404.
Vaithiyalingam, S., Warren, E.M., Eichman, B.F., Chazin, W.J., 2010. Insights into eukaryotic DNA priming from the structure and functional interactions of the 4Fe-4S cluster domain of human DNA primase. Proc. Natl. Acad. Sci. USA 107, 13684–13689.
van den Bosch, M., Bree, R.T., Lowndes, N.F., 2003. The MRN complex: coordinating

and mediating the response to broken chromosomes. EMBO Reports 4, 844–849.
Van Goethem, G., Dermaut, B., Löfgren, A., Martin, J.-J., Van Broeckhoven, C., 2001. Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. Nat. Genet. 28, 211–212.
van Loon, B., Markkanen, E., Hübscher, U., 2010. Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. DNA repair 9, 604–616.
Vanankeren, S.C., Murray, D., Meyn, R.E., 1988. Induction and Rejoining of γ-Ray-Induced DNA Single- and Double-Strand Breaks in Chinese Hamster AA8 Cells and in Two Radiosensitive Clones. Radiation research 116, 511–525.
Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., Walter, J.C., 2003. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes & development 17, 1894–1908.
Vidal, A.E., 2001. XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. The EMBO Journal 20, 6530–6539.
Vitale, S., Sperduto, R.D., Ferris, F.L., 2009. Increased Prevalence of Myopia in the United States Between 1971-1972 and 1999-2004. Arch Ophthalmol 127, 1632–1639.
Volker, M., Moné, M.J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J.H.J., van Driel, R., van Zeeland, A.A., Mullenders, L.H.F., 2001. Sequential Assembly of the Nucleotide Excision Repair Factors In Vivo. Mol. Cell 8, 213–224.
Waga, S., Stillman, B., 1994. Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro. Nature 369, 207–212.
Walker, J.R., Corpina, R.A., Goldberg, J., 2001. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair : Abstract : Nature. Nature 412, 607–614.
Walter, P., Klein, F., Lorentzen, E., Ilchmann, A., Klug, G., Evguenieva-Hackenberg, E., 2006. Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon Sulfolobus solfataricus. Molecular Microbiology 62, 1076–1089.
Wan, L., Lou, J., Xia, Y., Su, B., Liu, T., Cui, J., Sun, Y., Lou, H., Huang, J., 2013. hPrimpol1/CCDC111 is a human DNA primase-polymerase required for the maintenance of genome integrity. EMBO Reports 14, 1104–1112.
Wang, J.C., 1996. DNA topoisomerases. Annual Review of Biochemistry 65, 635–692.
Wang, J.C., 1998. Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. Quarterly reviews of biophysics 2, 107–144.
Wang, T.S., Hu, S.Z., Korn, D., 1984. DNA primase from KB cells. Characterization of a primase activity tightly associated with immunoaffinity-purified DNA polymerase-alpha. J. Biol. Chem. 259, 1854–1865.
Wang, T.S.F., 1991. Eukaryotic DNA polymerases. Annual Review of Biochemistry 60, 513–552.
Wang, Y., L, F.C., Kaguni, L.S., 1997. Accessory Subunit of Mitochondrial DNA Polymerase from Drosophila Embryos: Cloning, Molecular Analysis, and Association in the Native Enzyme. J. Biol. Chem. 272, 13640–13646.
Warren, E.M., Huang, H., Fanning, E., Chazin, W.J., Eichman, B.F., 2009. Physical interactions between Mcm10, DNA, and DNA polymerase alpha. J. Biol. Chem. 284, 24662–24672.
Washington, M.T., Johnson, R.E., Prakash, L., Prakash, S., 2004. Human DNA polymerase iota utilizes different nucleotide incorporation mechanisms dependent upon the template base. Mol. Cell. Biol. 24, 936–943.
Washington, M.T., Johnson, R.E., Prakash, S., Prakash, L., 1999. Fidelity and

processivity of Saccharomyces cerevisiae DNA polymerase eta. J. Biol. Chem. 274, 36835–36838.
Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191.
Watson, J.D., Crick, F.H., 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171, 737–738.
Weiner, B.E., Huang, H., Dattilo, B.M., Nilges, M.J., Fanning, E., Chazin, W.J., 2007. An iron-sulfur cluster in the C-terminal domain of the p58 subunit of human DNA primase. J. Biol. Chem. 282, 33444–33451.
Weller, G.R., Doherty, A.J., 2001. A family of DNA repair ligases in bacteria? FEBS Lett. 505, 340–342.
Weller, G.R., Kysela, B., Roy, R., Tonkin, L.M., Scanlan, E., Della, M., Devine, S.K., Day, J.P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K.M., Bowater, R.P., Jeggo, P.A., Jackson, S.P., Doherty, A.J., 2002. Identification of a DNA nonhomologous end-joining complex in bacteria. Science 297, 1686–1689.
Wenger, C.D., Phanstiel, D.H., Lee, M.V., Bailey, D.J., Coon, J.J., 2011. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. Proteomics 11, 1064–1074.
Wickner, W., Brutlag, D., Schekman, R., Kornberg, A., 1972. RNA Synthesis Initiates In Vitro Conversion of M13 DNA to Its Replicative Form. Proceedings of the … 69, 965–969.
Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, R.D., Stoneking, M., 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biological Journal of the Linnean Society 26, 375–400.
Wold, M.S., 1997. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annual Review of Biochemistry 66, 61–92.
Wold, M.S., Kelly, T., 1988. Purification and characterization of replication protein A, a cellular protein required for in vitro replication of simian virus 40 DNA. Proc. Natl. Acad. Sci. USA 85, 2523–2527.
Xu, X., Vaithiyalingam, S., Glick, G.G., Mordes, D.A., Chazin, W.J., Cortez, D., 2008. The Basic Cleft of RPA70N Binds Multiple Checkpoint Proteins, Including RAD9, To Regulate ATR Signaling. Mol. Cell. Biol. 28, 7345–7353.
Yamtich, J., Sweasy, J.B., 2010. DNA polymerase family X: function, structure, and cellular roles. Biochim. Biophys. Acta 1804, 1136–1150.
Yang, W., 2005. Portraits of a Y-family DNA polymerase. FEBS Lett. 579, 868–872.
Yang, X., Gregan, J., Lindner, K., Young, H., Kearsey, S.E., 2005. Nuclear distribution and chromatin association of DNA polymerase alpha-primase is affected by TEV protease cleavage of Cdc23 (Mcm10) in fission yeast. BMC Mol. Biol. 6, 13.
Yano, K.-I., Morotomi-Yano, K., Wang, S.-Y., Uematsu, N., Lee, K.-J., Asaithamby, A., Weterings, E., Chen, D.J., 2007. Ku recruits XLF to DNA double-strand breaks. EMBO Reports 9, 91–96.
Yuzhakov, A., Kelman, Z., Hurwitz, J., O'Donnell, M., 1999. Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme. The EMBO Journal 18, 6189–6199.
Zahurancik, W.J., Klein, S.J., Suo, Z., 2014. Significant contribution of the 3“→5” exonuclease activity to the high fidelity of nucleotide incorporation catalyzed by human DNA polymerase ϵ. Nucleic Acids Res. 42, 13853–13860.
Zamenhof, S., Brawerman, G., Chargaff, E., 1952. On the desoxypentose nucleic acids from several microorganisms. Biochim. Biophys. Acta 9, 402–405.

Zerbe, L.K., Kuchta, R.D., 2002. The p58 subunit of human DNA primase is important for primer initiation, elongation, and counting. Biochemistry 41, 4891–4900.

Zhang, H., Chatterjee, A., Singh, K.K., 2006. Saccharomyces cerevisiae polymerase zeta functions in mitochondria. Genetics 172, 2683–2688.

Zhang, Y., Wu, X., Rechkoblit, O., Geacintov, N.E., Taylor, J.-S., Wang, Z., 2002. Response of human REV1 to different DNA damage: preferential dCMP insertion opposite the lesion. Nucleic Acids Res. 30, 1630–1638.

Zhang, Y., Yuan, F., Wu, X., Taylor, J.S., Wang, Z., 2001. Response of human DNA polymerase iota to DNA lesions. Nucleic Acids Res. 29, 928–935.

Zhang, Y., Yuan, F., Wu, X., Wang, M., Rechkoblit, O., Taylor, J.S., Geacintov, N.E., Wang, Z., 2000. Error-free and error-prone lesion bypass by human DNA polymerase kappa in vitro. Nucleic Acids Res. 28, 4138–4146.

Zhao, F., Wu, J., Xue, A., Su, Y., Wang, X., Lu, X., Zhou, Z., Qu, J., Zhou, X., 2013. Exome sequencing reveals CCDC111 mutation associated with high myopia. Hum Genet 132, 913–921.

Zheng, L., Baumann, U., Reymond, J.-L., 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. 32, e115.

Zhu, H., Nandakumar, J., Aniukwu, J., Wang, L.K., Glickman, M.S., Lima, C.D., Shuman, S., 2006. Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D. Proc. Natl. Acad. Sci. USA 103, 1711–1716.

Zhu, J., Newlon, C.S., Huberman, J.A., 1992. Localization of a DNA replication origin and termination zone on chromosome III of Saccharomyces cerevisiae. Mol. Cell. Biol. 12, 4733–4741.

Zhu, W., Ukomadu, C., Jha, S., Senga, T., Dhar, S.K., Wohlschlegel, J.A., Nutt, L.K., Kornbluth, S., Dutta, A., 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. Genes & development 21, 2288–2299.

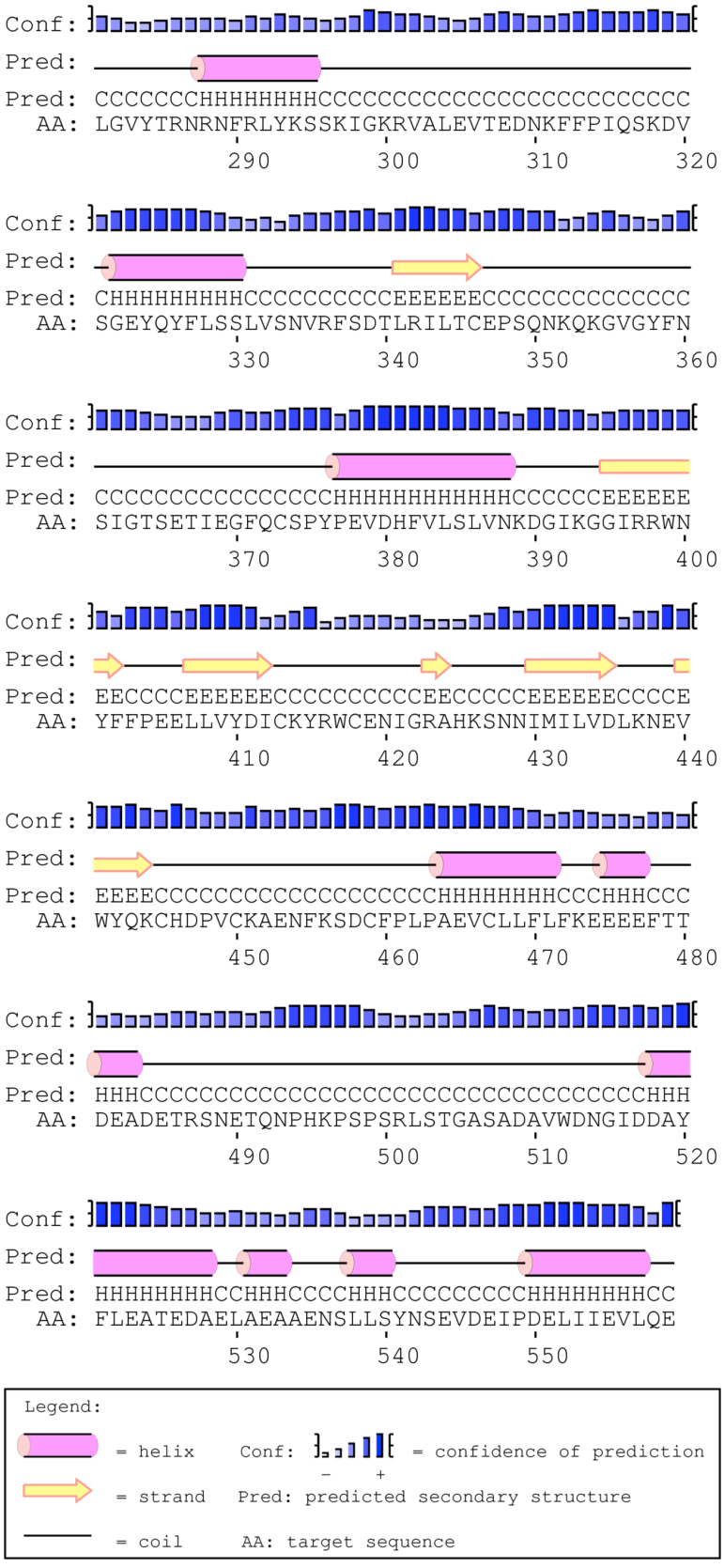
Zou, Y., Liu, Y., Wu, X., Shell, S.M., 2006. Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. J. Cell. Physiol. 208, 267–273.

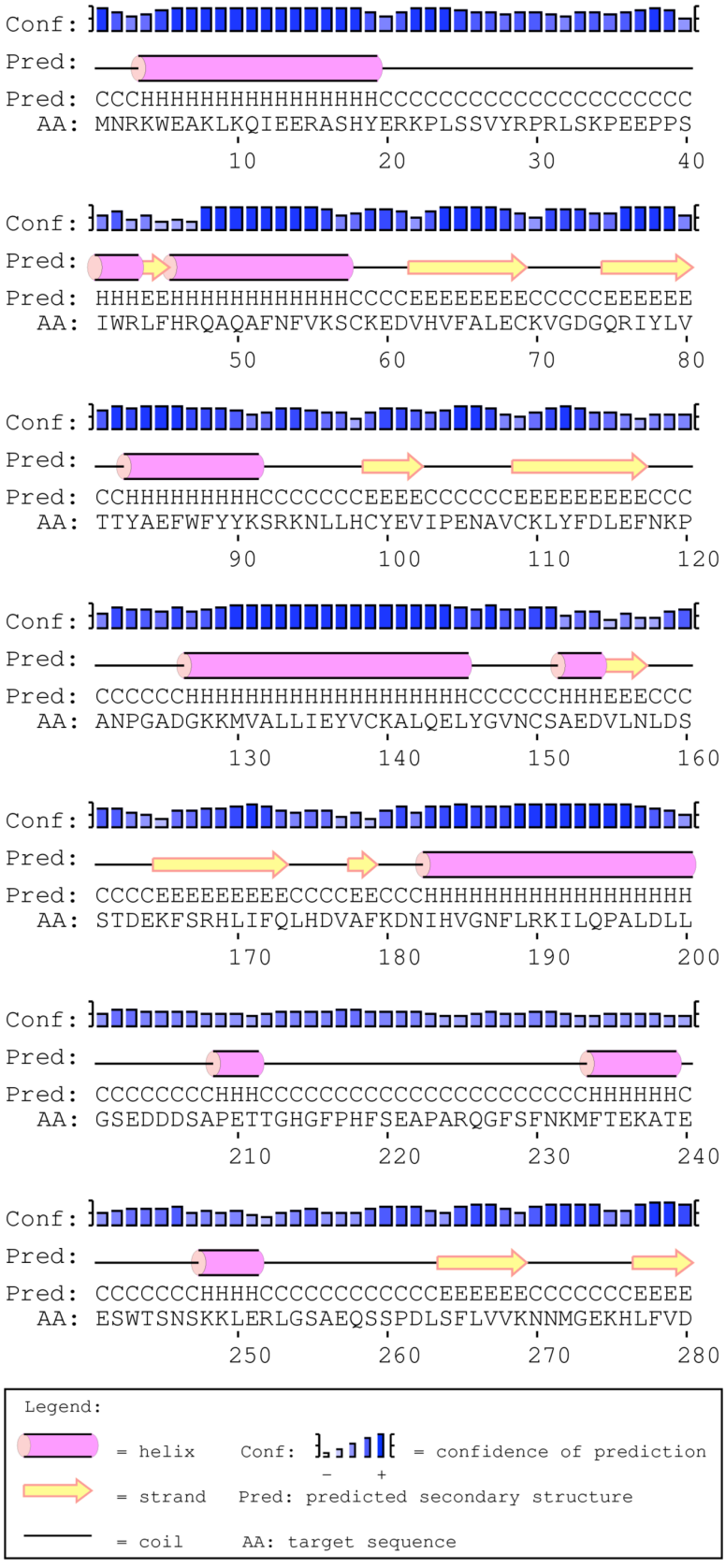
Zuo, Z., Rodgers, C.J., Mikheikin, A.L., Trakselis, M.A., 2010. Characterization of a Functional DnaG-Type Primase in Archaea: Implications for a Dual-Primase System. J. Mol. Biol. 397, 664–676.

Appendices

Appendix I

Secondary structure of human PrimPol from PSIPRED prediction.




Conf:
Pred:
Pred: CCCHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCC
AA: MNRKWEAKLKQIEERASHYERKPLSSVYRPRLSKPEEPPS
10 20 30 40
Conf:
Pred:
Pred: HHHEEHHHHHHHHHHHHCCCCEEEEEEEECCCCCEEEEEE
AA: IWRLFHRQAQAFNFVKSCKEDVHVFALECKVGDGQRIYLV
50 60 70 80
Conf:
Pred:
Pred: CCHHHHHHHHHCCCCCCCEEEECCCCCCEEEEEEEECCC
AA: TTYAEFWFYYKSRKNLLHCYEVIPENAVCKLYFDLEFNKP
90 100 110 120
Conf:
Pred:
Pred: CCCCCCHHHHHHHHHHHHHHHHHHHCCCCCCHHHEEECCC
AA: ANPGADGKKMVALLIEYVCKALQELYGVNCSAEDVLNLDS
130 140 150 160
Conf:
Pred:
Pred: CCCCEEEEEEEEECCCCEECCHHHHHHHHHHHHHHHHHH
AA: STDEKFSRHLIFQLHDVAFKDNIHVGNFLRKILQPALDLL
170 180 190 200
Conf:
Pred:
Pred: CCCCCCCCHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHC
AA: GSEDDDSAPETTGHGFPHFSEAPARQGFSFNKMFTEKATE
210 220 230 240
Conf:
Pred:
Pred: CCCCCCCHHHHCCCCCCCCCCCCEEEEEECCCCCCCEEEE
AA: ESWTSNSKKLERLGSAEQSSPDLSFLVVKNNMGEKHLFVD
250 260 270 280
Legend:
= helix
= strand
= coil
Conf: - +
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

Appendix II

Alignment of PrimPol from *X. tropicalis* and *X. laevis*.

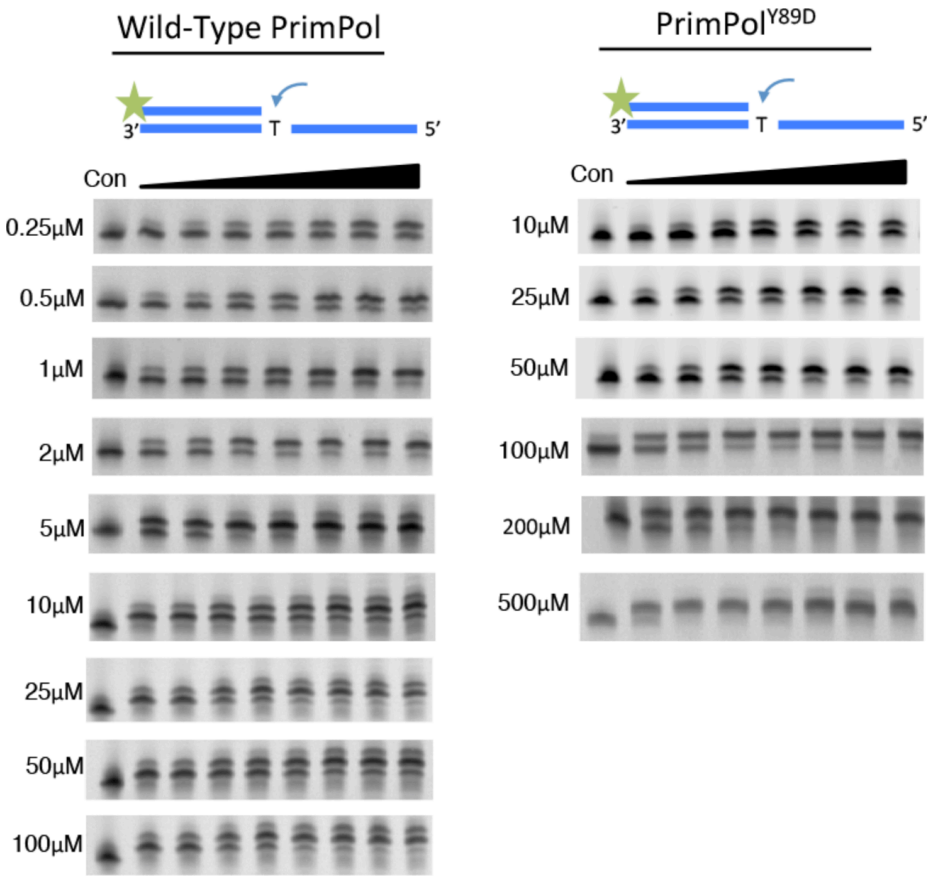
```
xlPrimpol       MRGKKTPESMERKWKEKVKNVENLADYYRRHPLSLIYRPKLAKPWLPSSVWKLFPRQALA 60
xtPrimpol       -----------------MKACGGLADYYRRHPLSFPYRPKLAKPWLPSSVWKLFPRQALA 43
                               :*   .***********: ************************

xlPrimpol       FNFAKTCKEDVHLFALETVIDDTERRLYLVTTYAEFWFYYVKHPFSLSHCYEVIPADAVC 120
xtPrimpol       FNFAKMCKEDVHVFALETVIDDTERRLYLVTTYAEFWFYYVKQPISLSHCYEVIPADTVC 103
                ***** ******:*******************************:*:************:**

xlPrimpol       KLYIDFEFYKSSNPEADGKKMVALVIEYFSIKLEELYGIKCSPEWVLNLDSSSKEKFSRH 180
xtPrimpol       KLYFDIEFYKPANPEVDGKKMVALVIEYFSTKLKEMYGIKCSPEWVLNLDSTTKEKFSRH 163
                ***:*:****.:***.************** **:*:***************::*******

xlPrimpol       LIFLLPNAAFKDNIHAGNFIKNVLKPLLPLAGYKSVTRDYVSGANESPCLPEAEVLCGNS 240
xtPrimpol       LIFVLPNAAFKDNIHAGNFIKHVLQPLLPLAGYKPVSQD-VFGANKSPCLPEAKNFCDAS 222
                ***:*****************:**:*********.*::* * ***:*******: :*. *

xlPrimpol       GKPSLAKAVKDVQSIENCDLSSLIVRDKYGGNQLCIDLGVYTKNRNFRLYKASKLGKNVP 300
xtPrimpol       RKPSQANAAKHVHSIENGDLSSLIVRDKYGGSQLLIDLGVYTKNRNFRLYKSSKLGKNVP 282
                 *** *:*.*.*:**** *************.** ****************:********

xlPrimpol       FTLAEDNKFRSKPQKDICTEEHIFLCSLISNIRFSDSLKILTCSSPDYGMNKATSPYNKN 360
xtPrimpol       FMLAEDNKFRSKPPKDFSAEEHIFLCSLISNVRFSDSLKILTCSNPDYRTSKAASHNNKN 342
                * *********** **:.:************:************.***  .**:*  ***

xlPrimpol       PGVTIKGYQFSPYPEIDDFILSLVTQEGFQGSIRSWNYFSSEELLVYETANYRWCANIGR 420
xtPrimpol       PGVTMKGFQVSPYPEIDDFILSLVTREGFQGSIRSWHYFSSEELLVYETANYRWCANIRR 402
                ****:**:*.***************:**********:******************** *

xlPrimpol       AHKSNNVMLLVDLKREVWYQKCHDPICRAQNYKSECYPLPPEVCLPFLFKEEDEESVFTM 480
xtPrimpol       AHKSNNVMLLVDVKREVWYQKCHDPICRAQNFKSECLPLPSEVCLPFLFKEDDEESIFTM 462
                ************:******************:**** ***.**********:****:***

xlPrimpol       DENGNIKETKINRCHVLFADGYLPPQKPSMPRSGNTMASGERHSGSEIDDACILEATEDV 540
xtPrimpol       DENGNIKETKINRSLVPMAD------KPSIARSGETMGPWERHCDSEIDDACILEAIEDI 516
                *************. * :**      ***:.***:**.. ***..*********** **:

xlPrimpol       EFVNAVDTSLAPLDLDDVQFSGEDVEFANAVDNSLAHLDLDDVQFSGEDIVFANAVDTSL 600
xtPrimpol       EFSDAVDTSLAHLELDDVGIPDEDIAFVNAVDTSLAHLELDYVGIPDEDIVFVNAVDTSL 576
                ** :******* *:**** :..**: *.****.*****:** * :..*****.*******

xlPrimpol       AHLNLE---IPDEDIEFANAVDTSLAHLDSEDVEIPDE---------------------- 635
xtPrimpol       AHLELDDVGIPDEDIAFVNAVDTSLAHLGSDDVGIPDEDIEFVNAVDNSLAHLDSDDLKI 636
                ***:*:   ****** *.**********.*:** ****

xlPrimpol       ----------------------DIEFANAVDTSLAHLDLDDAEIPDTLLLESLYKHEMF 672
xtPrimpol       SDEDVAFVNAVDNSLAHLDSDDEDIEFVNAVDTSLAHLDSDDVEIPDELLLKSLYEHEMF 696
                                      ****.*********** **.**** ***:***:****

xlPrimpol       ASK 675
xtPrimpol       ASK 699
                ***
```

xlPrimPol = *X. laevis* PrimPol

xtPrimPol = *X. tropicalis* PrimPol

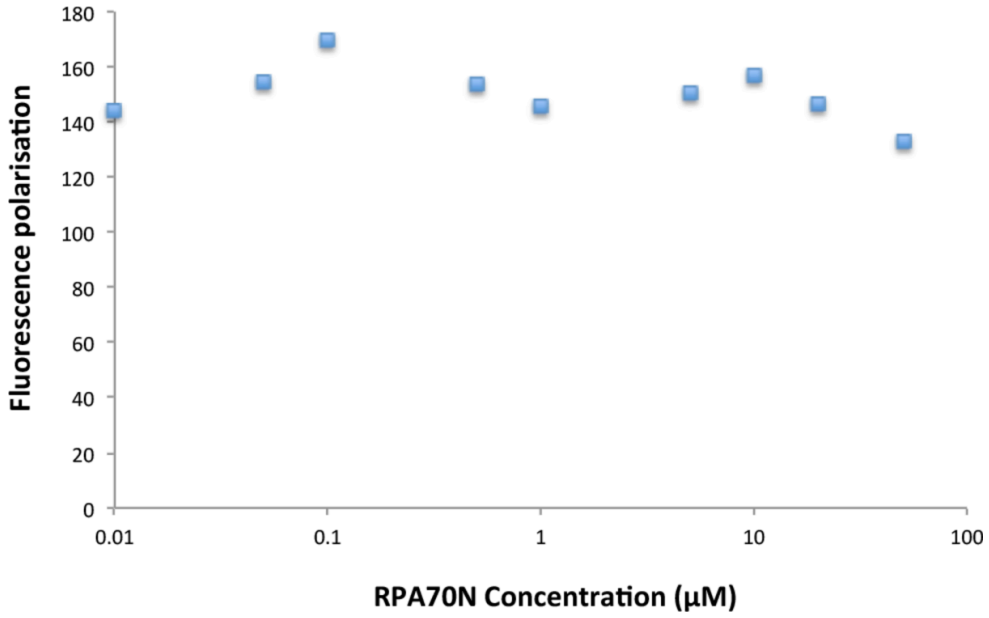
Appendix III

Representative single incorporation extension assays for each of the concentrations of dATP tested (Figures 4.2 and 5.7) showing extension relative to the control bands in the first lane. Each of these gels were repeated in triplicate.



Appendix IV

Testing the interaction of RPA70N with a fluorescently labelled RBM-A peptide by fluorescence anisotropy. The peptide was subsequently confirmed as unlabelled by mass spectrometry.



Nucleic Acids Research Advance Access published March 27, 2014
Nucleic Acids Research, 2014 **1**
doi: 10.1093/nar/gku214

Molecular dissection of the domain architecture and catalytic activities of human PrimPol

Benjamin A. Keen†, Stanislaw K. Jozwiakowski†, Laura J. Bailey, Julie Bianchi and Aidan J. Doherty*

Genome Damage and Stability Centre, University of Sussex, Brighton, BN1 9RQ, UK

Received November 8, 2013; Revised February 28, 2014; Accepted February 28, 2014

ABSTRACT

PrimPol is a primase–polymerase involved in nuclear and mitochondrial DNA replication in eukaryotic cells. Although PrimPol is predicted to possess an archaeo-eukaryotic primase and a UL52-like zinc finger domain, the role of these domains has not been established. Here, we report that the proposed zinc finger domain of human PrimPol binds zinc ions and is essential for maintaining primase activity. Although apparently dispensable for its polymerase activity, the zinc finger also regulates the processivity and fidelity of PrimPol's extension activities. When the zinc finger is disrupted, PrimPol becomes more promutagenic, has an altered translesion synthesis spectrum and is capable of faithfully bypassing cyclobutane pyrimidine dimer photolesions. PrimPol's polymerase domain binds to both single- and double-stranded DNA, whilst the zinc finger domain binds only to single-stranded DNA. We additionally report that although PrimPol's primase activity is required to restore wild-type replication fork rates in irradiated *PrimPol*$^{-/-}$ cells, polymerase activity is sufficient to maintain regular replisome progression in unperturbed cells. Together, these findings provide the first analysis of the molecular architecture of PrimPol, describing the activities associated with, and interplay between, its functional domains and defining the requirement for its primase and polymerase activities during nuclear DNA replication.

INTRODUCTION

DNA replication is an essential biological process, indispensable for the existence of life. DNA replication systems rely on a semi-conservative mode of replication where the initiation of DNA synthesis requires a free 3′ hydroxyl group to which additional nucleotides are subsequently added by replicative polymerases (1). Genome replication starts with DNA template-dependent synthesis of short RNA primers that are further extended with deoxynucleotides by the replication machinery. This initial step in DNA replication is often defined as *de novo* primer synthesis and is catalysed by specialised DNA polymerases known as primases. Based on their structural topology, these enzymes can be classified into archaeo-eukaryotic primases (AEPs) or DnaG-like prokaryotic primases (2,3).

Until recently, the Pol α-associated DNA primase small subunit (PriS) that is responsible for *de novo* polymer synthesis through the production of RNA primers was considered to be the sole AEP superfamily member present in eukaryotes (1). However, bioinformatic analysis identified the existence of an additional uncharacterized DNA primase in eukaryotes called PrimPol (CCDC111 or FLJ33167) (3–7), which belongs to the 'NCLDV-herpesvirus clade' of viral AEPs.

Recent studies have reported that PrimPol is a DNA primase (4–7), with the ability to synthesise primers using either ribonucleotides (NTPs) or deoxyribonucleotides (dNTPs), preferring to make DNA primers. In addition, the enzyme possesses robust template-dependent DNA polymerase activity (4–7). PrimPol is present in both the nucleus (4–6) and mitochondria (7) of eukaryotic cells. The enzyme localises to nuclear chromatin during replication (4–6) and this recruitment is more pronounced after treatment with damaging agents (e.g. ultraviolet light; UV) or replication stalling drugs (e.g. hydroxyurea) (4). UV irradiation can induce the covalent linkage of adjacent pyrimidines leading to the formation of cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6–4) pyrimidone photoproducts ((6–4)PPs). These helix-distorting lesions interfere with major biological processes, including DNA replication and transcription (8). PrimPol can perform translesion synthesis (TLS) bypass of the highly distorting (6–4)PPs but

*To whom correspondence should be addressed. Tel: +44 1723 877500; Email: ajd21@sussex.ac.uk
†Contributed equally to this work.
Present address:
Julie Bianchi, Department of Oncology-Pathology, Cancer Center Karolinska (CCK), Karolinska Institutet, SE-171 76, Stockholm, Sweden.
The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

is also involved in replication through oxidative lesions, including 8-oxoguanine (8-oxoG) (4,5,7). Deletion of PrimPol (*PrimPol*$^{-/-}$) induced replication fork slowing, which was much more pronounced when cells were UV irradiated (4). Knockout cells also exhibited increased formation of chromosomal breaks, particularly after aphidicolin treatment. This study also reported that PrimPol is not epistatic with the Pol η-dependant CPD-bypass pathway and therefore appears to form an independent pathway required for the bypass of UV, and other lesions, during replication (4,5).

Comparative analysis of the amino acid sequence of eukaryotic PrimPols identified the presence of two distinctive domains (3), an enzymatic AEP polymerase and a UL52-like zinc finger (Zfn) domain (Figure 1A). Mutation of the conserved zinc-chelating cysteine residues in the UL52 domain of herpes simplex virus type 1 resulted in a severe reduction in DNA binding (8,9). The AEP polymerase domain contains the three signature catalytic motifs. A highly conserved DxE motif (motif I), together with an aspartic acid residue from motif III, forms the divalent metal binding site. Motif II (SxH) forms part of the putative nucleotide binding motif. Mutagenesis of these motifs abolishes the catalytic activities of these polymerases (4,5,7,10,11).

In this study, we have dissected the molecular architecture of human PrimPol to define the activities associated with its two major functional domains. We demonstrate that the zinc finger domain is crucial for its primase activity. PrimPol also has DNA template-dependent DNA polymerase activities and this bifunctional enzymatic activity is reminiscent of archaeal replicative primases (11–14). Although PrimPol's polymerase activities appeared, initially, to be largely independent of the zinc finger domain, our data suggests that the zinc finger domain plays important roles in the processivity and fidelity of DNA synthesis. Additionally, we demonstrate that the zinc finger domain has regulatory roles for both DNA/RNA primase and TLS activities. We also report that a catalytically active fragment of human PrimPol, containing only the AEP domain (PrimPol$_{1–354}$), catalyzes TLS bypass of both major UV-induced DNA lesions, CPDs and (6–4)PPs. Analysis of the DNA binding affinities of catalytically active PrimPol$_{1–354}$ and the C-terminal UL52-like domains established that the enzymatic domain can bind both single-stranded (ss) and double-stranded (ds) DNA, whilst the zinc finger domain can only bind to ss DNA. Finally, we report that although PrimPol's polymerase activity is sufficient to maintain wild-type replication fork rates in unperturbed *PrimPol*$^{-/-}$ cells, primase activity is requisite for normal replication in UV-treated cells. Together, these findings provide the first molecular insights into the domain architecture of human PrimPol, defining the activities associated with its major functional modules and delineating the requirement for its specific catalytic activities during perturbed and unperturbed DNA replication.

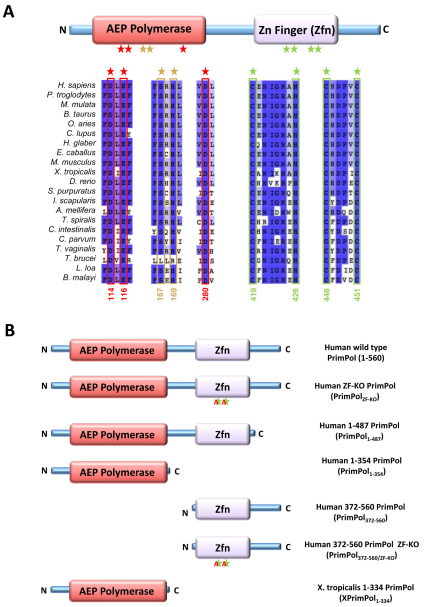


Figure 1. Domain architecture of eukaryotic PrimPol. (**A**) PrimPol is composed of an AEP primase–polymerase and a Zn^{2+} finger (Zfn) domain. The AEP domain contains three conserved catalytic motifs (I–III). The first motif (motif I, the first two red stars) is the DxE motif that, along with the conserved D in the third motif (motif III, third red star), forms the catalytic triad that coordinates divalent metal ions essential in the synthesis of oligonucleotide chains. The second motif (motif II, two light-orange stars) is a conserved SxH motif that is required for the coordination of the incoming nucleotide. The zinc finger domain contains a canonical C–H–C–C motif that coordinates a zinc ion and stabilises the anti-parallel β-sheet and α helix structure. The numbers below the conserved residues denote their positions in human PrimPol. (**B**) A number of human PrimPol constructs were produced to analyse the domain architecture of PrimPol. A Zfn knockout mutant was produced by mutating the first conserved cysteine and histidine residues that coordinate the zinc ion (C419A and H426A, respectively). We constructed a number of truncation mutants based on secondary structure predictions. PrimPol$_{1–487}$ lacks the C-terminus region downstream of the Zfn domain. PrimPol$_{1–354}$ deletion mutant contains the polymerase domain but lacks the Zfn domain. PrimPol$_{372–560}$ possesses the zinc finger domain but lacks the polymerase domain. PrimPol$_{1–354/ZF-KO}$ contains only the zinc finger domain but the C419A and H426A are mutated. A 1–334 deletion mutant in *Xenopus tropicalis* (XPrimPol$_{1–334}$) was additionally constructed that is equivalent to PrimPol$_{1–354}$.

MATERIALS AND METHODS

Construction of human PrimPol mutants

Human PrimPol was cloned as described previously (4). A number of PrimPol mutants were constructed by polymerase chain reaction (Figure 1B) and the primers used can be found in the supplementary data (Supplementary Table S1). *Xenopus tropicalis* PrimPol cDNA was sub-cloned into pET28a (Novagen) using BamHI and XhoI restriction sites. Subsequently, a 1–334 C-terminal deletion (XPrimPol$_{1–334}$) mutant was also constructed (primers in Supplementary Table S1). Additionally, we constructed a PrimPol$_{AXA}$ (cat-

alytically null) mutant that was cloned and purified, as previously described (4).

Expression and purification of recombinant PrimPol proteins

Wild-type PrimPol and $PrimPol_{1–487}$ were expressed in *Escherichia coli* BL21(pLysS) cells overnight at 16°C, as described previously (4). $PrimPol_{1–354}$, $PrimPol_{372–560}$ and $XPrimPol_{1–334}$ were expressed overnight at 25°C. All of the proteins were then purified by the same protocol, except where otherwise noted. Cells were pelleted and resuspended in buffer A (50 mM Tris–HCl (pH 7.5), 200 mM NaCl, 30 mM imidazole, 10% (v/v) glycerol, 17 μg/ml PMSF, 34 μg/ml benzamidine) supplemented with 0.5% IGEPAL. Cells were disrupted by sonication and proteins isolated by centrifugation. All proteins were purified by affinity chromatography using a 25 ml Ni^{2+}-NTA agarose (Qiagen) column equilibrated with buffer A and eluted with buffer B (as A, with 300 mM imidazole). The eluted protein-containing fractions were then diluted 1 in 10 in buffer C (50 mM Tris–HCl (pH 7.5), 10% (v/v) glycerol) and separated by charge by affinity exchange chromatography on a 5 ml HiTrap Heparin HP column (GE Healthcare) equilibrated with buffer C. $PrimPol_{372–560}$ was purified on a 5 ml HiTrap Q HP (GE Healthcare) column equilibrated with buffer C. All proteins were subject to gradient elution with buffer D (as C, with 2M NaCl) and protein-containing fractions were purified by size exclusion chromatography on a Superdex S-75 analytical gel-filtration column (GE Healthcare) equilibrated in buffer E (50 mM Tris–HCl (pH 7.5), 300 mM NaCl, 10% (v/v) glycerol). Protein concentrations were determined by absorption spectra at 280 nm.

DNA primase assays

The sequences of the oligonucleotides used in the primase assays are sequences 1–4 in Supplementary Table S2. The non-radioactive primase assay was performed in three subsequent steps. Typically detection of primase activity was started from incubation of 1 μM of the enzyme to be tested in 20 μl reaction volume containing 500 nM homopolymeric ss DNA templates with a biotin modification at the 5′ end (see sequences 1–4 in Supplementary Table S2), 500 μM rNTPs (Invitrogen) or 500 μM dNTPs (Roche), 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$, 50 mM NaCl. This primer synthesis reaction was carried out for 2 h at 37°C. Following the primer synthesis reaction, the reaction mixture was supplemented with 0.2 U of Klenow Taq (purified as in Engelke *et al.* (15)) and 15 μM FAM-6-dATP (Jena-Biosciences), incubated at 37°C for 45 min to allow fluorescent labelling of *de novo* synthesised primers. The primer synthesis/labelling enzymatic reactions were terminated by adding 450 μl of binding-washing (B-W) buffer (10 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM EDTA). The quenched reactions were subsequently added to 30 μl of streptavidin-coated beads (Invitrogen) and mixed on a spinning wheel for 1 h at 4°C. After binding the ss DNA templates, the suspensions were spun down briefly to sediment the beads. The supernatant was removed and the beads were washed three times with 1 ml B-W buffer. The beads were then suspended in 20 μl of the B-W buffer supplemented with equal volume of loading buffer (8 M Urea, 10 mM EDTA). The resulting sample was incubated at 95°C for 3 min in order to liberate the primers synthesised *de novo*. Subsequently, the samples were spun down briefly and 20 μl volume of each reaction was resolved by standard denaturing electrophoresis using a 15% (v/v) polyacrylamide gel matrix containing 7 M urea and 1× TBE buffer (100mM Tris, 100mM Boric Acid, 2mM EDTA). Typically the electrophoretic separation of primase assay products was performed at 850 V for ∼2 h. After electrophoresis gels were scanned for fluorescent signal detected by a Fujifilm FLA-5100 image reader.

DNA primer extension assays

Hex-labelled DNA primers were annealed to oligomer templates (sequences in Supplementary Table S2). 34 nM of protein was incubated with 20 nM DNA, 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$, 1 mM DTT and 200 μM dNTPs (Roche) to a final volume of 20 μl. In the case of single nucleotide incorporations 200 μM of each individual nucleotide adenosine, cytosine, guanosine or thymidine triphosphate (dATP, dCTP, dGTP or dTTP respectively) were added in lieu of dNTPs. The reactions were terminated by the addition of 2× stop buffer (95% formamide, 0.09% xylene cyanol, 0.05% bromophenol blue, 200 nM competitor oligonucleotide) and boiled at 95°C for 5 min. Samples were resolved by electrophoresis on a 15% (v/v) polyacrylamide gel containing 7 M urea and 1× TBE buffer at 850 V for 2.5 h in 1× TBE buffer. Fluorescently labelled DNA oligomers were detected by scanning using a Fujifilm FLA-5100 image reader.

Terminal transferase assays

The terminal transferase capability of wild-type PrimPol and its variants was studied using three types of synthetic DNA substrates: ds DNA (sequence 6 annealed to sequence 7 from Supplementary Table S2) and a primer template containing both single-stranded and double-stranded DNA interfaces (sequence 5 annealed to sequence 6; Supplementary Table S2). Typical reactions were performed in 20 μl volume containing 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM NaCl, 10 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM DTT, 20 nM DNA substrate, 200 μM dNTPs (Roche), with 100 nM recombinant human PrimPol or its variants. All reactions were incubated at 37°C for 30 min, reactions were quenched with addition of 2× stop buffer (95% formamide, 0.09% xylene cyanol, 0.05% bromophenol blue, 200 nM competitor oligonucleotide) and boiled at 95°C for 5 min. Samples were resolved by electrophoresis as described for the primer extension assays.

DNA extension processivity assays

Full-length and $PrimPol_{1–354}$ constructs were analysed for processivity as described previously (16). PrimPol (100 nM) was preincubated at 37°C with 60 nM substrate DNA (sequence 5 annealed to sequence 6 from Supplementary Table S2), 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$ and 1 mM DTT for 30 min. Reactions were initiated by adding dNTPs and excess of sonicated herring

sperm DNA (1 mg/ml) as an enzyme trap. Reactions were quenched after time points of 15, 30, 60, 120 and 360s with addition of 2× stop buffer (95% formamide, 0.09% xylene cyanol, 0.05% bromophenolblue, 200 nM competitor oligonucleotide) and boiled at 95°C for 5 min. Samples were resolved by electrophoresis as described for the primer extension assays. To test the effectiveness of the trap, each of the enzymes were also preincubated with excess of herring sperm DNA (1 mg/ml), as well as 40 nM DNA, 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$ and 1 mM DTT. To ensure the trap did not disrupt the processivity of the enzyme, the experiment was also carried out with the Klenow fragment of Taq polymerase.

The band intensities at 360 s were measured using ImageQuant software (GE Healthcare). The percentage of active polymerases at a given position is given by Equation (1):

$$\%\ \text{active polymerases at } n = (I_n + I_{n+1} + I_{n+2\ldots}) \times 100\%/(I_1 + I_2 + I_3 \ldots) \quad (1)$$

Where I_1 is the intensity at position 1, I_n is the intensity at position n and so on.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSAs) were carried out on two synthetic substrates: ss DNA (sequence 7 from Supplementary Table S2) and ds DNA (sequence 6 annealed to sequence 7 from Supplementary Table S2). Varying concentrations of the human $PrimPol_{1–354}$ and $PrimPol_{372–560}$ constructs were added to 40 nM DNA, 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$ and 1.0 mM DTT to a final volume of 20 μl and incubated at 25°C for 60 min. The reactions were supplemented with 2 μl 25% (w/v) ficoll. Samples were resolved by electrophoresis on a 5% (v/v) polyacrylamide gel containing 0.5× TBE buffer at 150 V for 0.5 h, then 300 V for 2.5 h in 0.5× TBE buffer. Fluorescently labelled DNA oligomers were detected by scanning using a Fujifilm FLA-5100 image reader.

Inductively coupled plasma mass spectrometry

The concentrations of zinc ions in protein samples were determined using inductively coupled plasma mass spectrometry (ICP-MS). For this analysis of human PrimPol, the proteins were gel filtered in zinc-free solutions to minimise the concentration of zinc ions that were not chelated by protein. Quantitative calibrations were made using standards prepared from zinc solutions containing 0, 100, 500 and 1000 ng/ml. Each sample was measured in triplicate and background measurements of buffer without protein were determined and subtracted from the appropriate readings. A 2% nitric acid wash was performed between standards and samples.

PrimPol protein denaturation assays

The full-length and $PrimPol_{1–487}$ constructs were both dialysed into a buffer of 200 mM NaF, 15 mM Tris-HCl (pH 7.5). 45 μl of 1 μM protein solution was added to 15 μl of SYPRO Orange, resulting in a final protein concentration of 0.75 μM. A control of 45 μl buffer was also added to 15 μl of SYPRO Orange. 20 μl of each sample was aliquoted in triplicate into a 96-well plate. Protein melting experiments were carried out using the LightCycler 480 System II (Roche). The instrument was configured with a detection format of 465 nm as the wavelength of excitation and 580 nm as the emission wavelength to detect SYPRO Orange-specific signal. Denaturation curve fluorescent signal was acquired within a range of 20–80°C using a ramping rate of 0.03°C s^{-1} and an acquisition of 20 data points per degree celsius. Melting temperatures (T_m) were determined through the measurement of the lowest point of the negative differential of the denaturation curve. Data was corrected for the background signal of the buffer conditions and presented as units of fluorescence with respect to temperature ± 1 SD.

Complementation and survival assays in *PrimPol*$^{-/-}$ avian cells

DT40 cells were grown at 39°C in RPMI 1640 medium supplemented with 10 μM β-mercaptoethanol, penicillin, streptomycin, 10% foetal calf serum and 1% chicken serum (Sigma). DT40 cells deleted of PrimPol made previously (4) were stably complemented with wild-type and truncated or mutated forms of human PrimPol cloned into the pCI-neo vector by electroporation as described previously (4). Stable expressing cells were selected using 2 mg/ml G418 (Sigma) and confirmed by western blot. DNA fibre analysis was carried out as described previously (4). Fibre analysis was carried out in triplicate ($n = 3$).

RESULTS

Zinc finger domain of PrimPol coordinates a zinc ion

The C-terminal domain of human PrimPol contains a highly conserved UL52-like Cys-His-Cys-Cys (CHC2) motif, predicted to fold into a functional zinc finger (Zfn) motif. (Figure 1A) To establish if this is a *bona fide* zinc-binding motif, we produced C-terminal truncations in PrimPol containing either an intact ($PrimPol_{372–560}$) or mutated ($PrimPol_{372–560/ZF-KO}$) Zfn motif. $PrimPol_{372–560/ZF-KO}$ has the first two residues of the Zfn motif mutated to alanine. The proteins were gel filtered in the absence of zinc and concentrations of zinc were then measured using ICP-MS (Supplementary Table S3). The ICP-MS analysis revealed that the concentration of the zinc in the UL52-like domain represented an occupancy of 80.38%, suggesting that a small fraction of the protein is either unfolded or has another divalent metal ion chelated. In contrast, the zinc finger knockout had an occupancy of 3.90%, suggesting that the first two cysteine and histidine residues of this motif are required for the coordination of zinc. We also measured the zinc occupancy in the polymerase domain ($PrimPol_{1–354}$) after purification in the absence of zinc and the occupancy in this case was 1.57%. The detection of low levels of zinc in the polymerase domain is to be expected as it contains divalent cation-binding motifs. These motifs generally bind magnesium or manganese, which have respective ionic radii of 86 and 81 pm, but could conceivably bind zinc, with a comparable ionic radius of 88 pm (17).

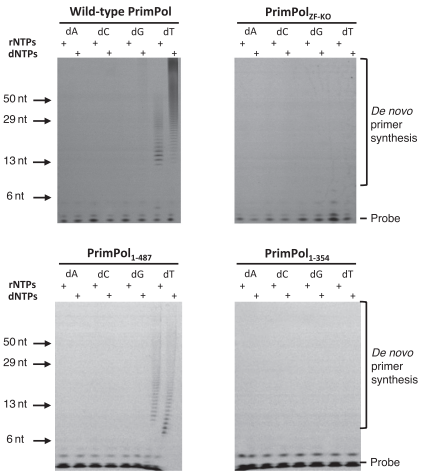


Figure 2. Primase activity of human PrimPol. (**A**) Human PrimPol has primase activity and can produce *de novo* primers using rNTPs and dNTPs opposite a poly(dT) template. (**B**) PrimPol$_{ZF-KO}$ lacks *de novo* primer synthesis activity, suggesting that an intact zinc finger is required for primase activity. (**C**) PrimPol$_{1-487}$ also has primase activity similar to the wild-type PrimPol. The unstructured region that is downstream of the zinc finger is therefore not required for primase activity. (**D**) PrimPol$_{1-354}$ has no primase activity, which indicates that PrimPol requires a functional zinc finger for primer synthesis.

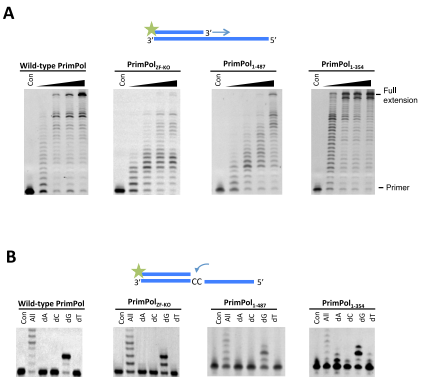


Figure 3. Polymerase activity and fidelity of human PrimPol. (**A**) Human PrimPol was incubated with dNTPs and substrate at 1, 3, 5 and 30 min time points. PrimPol was proficient at extending an undamaged oligonucleotide template using dNTPs. Human PrimPol did not require an intact zinc finger in order to carry out primer extension, as evidenced by the extension of primers by PrimPol$_{ZF-KO}$ and PrimPol$_{1-354}$. PrimPol$_{1-487}$ that lacked the unstructured C-terminus of the protein was also polymerase proficient. PrimPol$_{1-354}$ exhibited a higher rate of polymerase activity compared to the other constructs. (**B**) Incorporation of nucleotides opposite two templating cytosine bases. PrimPol was incubated for 5 min with the DNA substrate and each of the dNTPs. All four of these PrimPol constructs inserted two guanine nucleotides opposite two cytosines in Watson–Crick base-pairing manner. PrimPol$_{1-354}$ could additionally incorporate a single adenine opposite the first cytosine.

PrimPol's zinc finger domain is essential for primase activity

Previously, we and others have shown that PrimPol has DNA primase activity *in vitro* (4–7), preferentially catalysing synthesis of primers composed of DNA. Previous studies on the viral HSV1 helicase/primase complex showed that mutation of key residues in the UL52 zinc-binding domain of this enzyme resulted in loss of primase activity (8,9). Alignment of human PrimPol with other eukaryotic orthologues identified the conserved residues (C419, H426, C446 and C451; Figure 1A) predicted to coordinate a zinc ion. We prepared a zinc finger knockout enzyme carrying C419A and H426A point mutations (PrimPol$_{ZF-KO}$) (Figure 1B).

As previously observed (4,5), wild-type PrimPol exhibited primase activity on a 60-mer poly-dT ss DNA template, however, PrimPol$_{ZF-KO}$ showed no detectable primase activity (Figure 2). Furthermore, we analysed the primase activity of two deletion mutants, PrimPol$_{1-487}$ and PrimPol$_{1-354}$, which were constructed to determine whether elimination of primase activity resulted from a structural change brought about by the point mutations or whether the zinc-finger domain is necessary for primase activity. As expected, a deletion mutant lacking the entire zinc finger (PrimPol$_{1-354}$) showed no detectable primase activity (Figure 2). However, PrimPol$_{1-487}$, containing the entire zinc-finger domain but lacking the smaller C-terminal portion, predicted to be disordered, was an active DNA primase (Figure 2). Notably, the observed primase activity of PrimPol$_{1-487}$ was lower than the activity observed for the parental enzyme, possibly as this truncation results in partial loss of UL52 domain functionality. Together, our data provide experimental evidence that the zinc-finger domain, but not the disordered region proximal to this domain, is essential for the primase activity of human PrimPol.

PrimPol's DNA polymerase activity is modulated by the zinc finger domain

As the zinc-finger domain is required for the priming activity of human PrimPol, we also evaluated the importance of this functional module in the context of PrimPol's DNA polymerase activity. To address this, we employed standard primer extension assays using unmodified synthetic primer templates and dNTPs. We compared the relative DNA polymerase activities of four different variants of human PrimPol including the wild-type enzyme, PrimPol$_{ZF-KO}$, PrimPol$_{1-487}$ and PrimPol$_{1-354}$ (Figure 1B; Table 1A). Although all tested variants of PrimPol were proficient at DNA template-dependent DNA synthesis, we observed variation in the relative specific polymerase activity, processivity and fidelity among these enzymes (Figure 3A). PrimPol$_{ZF-KO}$ showed a pronounced decrease in specific

DNA polymerase activity. In contrast, PrimPol$_{1–354}$ exhibited both increased relative DNA processivity and extension activities. Despite variations in specific DNA polymerase activities across the variants of human PrimPol tested, the zinc-finger module does not appear to be essential for DNA polymerase activity of this enzyme. However, a regulatory role in catalysis for this functional module cannot be excluded.

Consequently, we next examined the potential regulatory role of this functional module. In particular, we tested the fidelity of incorporation of each of the four single deoxynucleotides opposite two templating cytosines (Figure 3B). Wild-type PrimPol, PrimPol$_{ZF-KO}$ and PrimPol$_{1–487}$ all incorporated two correct incoming guanine nucleotides opposite the templating cytosines at the N+1 and N+2 positions. In contrast, PrimPol$_{1–354}$ that lacks the entire zinc finger domain showed a significantly reduced ability to accurately select the incoming nucleotide. In particular, we observed significant mis-incorporation of adenine nucleotides opposite templating cytosine at the N+1 position and mis-incorporation of another guanine nucleotide opposite a templating adenine at N+3. This observation suggests the potential importance of the Zfn domain in influencing the fidelity of PrimPol's polymerase activity. PrimPol cannot incorporate NTPs during primer extension *in vitro* and mutation of the Zfn domain did not alter this preference (unpublished data). This is in contrast to the AEP domain of LigD, which can extend primers using both dNTPs and NTPs (18,19).

PrimPol exhibits template-independent primer extension activity in the presence of manganese

The flexible extension activities reported previously for Non homologous end-joining (NHEJ) AEP polymerases include limited template-independent terminal transferase activity on both ss DNA and blunt-ended ds DNA (18,19). Such activities are often stimulated by the presence of manganese ions, which can accelerate the rate of DNA synthesis and, as a result, lower the overall fidelity of this process. We therefore tested if human PrimPol also possesses terminal transferase activity in the presence of manganese. Wild type, as well as all three variants, of human PrimPol showed no detectable terminal transferase activity when tested on a blunt ended ds DNA substrate (Figure 4, left panels). However, PrimPol's DNA synthesis activity appeared to be significantly increased when primer extension assays were performed in the presence of dNTPs (all) and manganese (Figure 4A, right panel). To investigate if this increased activity resulted from altered fidelity, we repeated the assays with each of the dNTPs and observed that PrimPol catalysed multiple extension of the primer even in the presence of a single nucleotide pool. This promiscuous DNA synthesis was most efficient when enzyme was utilising dATP or dTTP. In both cases, extension of primers by up to ∼20 nucleotides was observed. Notably, this phenomenon was not observed when magnesium was used instead of manganese. When a single incorporation of dCTP was investigated, we observed significantly slower extension rates but the enzyme was able to extend the primer to produce predicted full-length product. In contrast, incorporation of the correct incoming nucleotide (dGTP) resulted in least pronounced primer extensions up to 16 nucleotides.

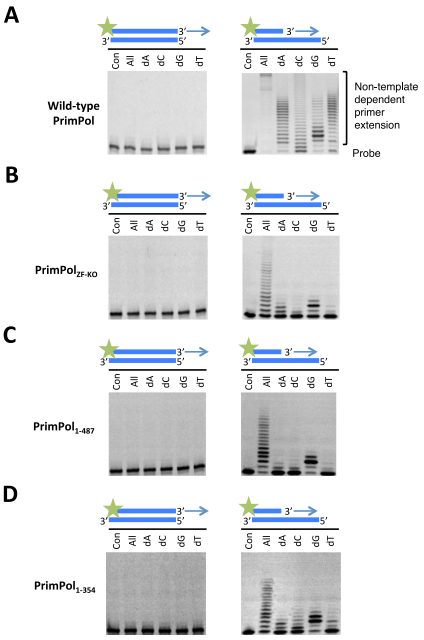


Figure 4. Template-independent extension in the presence of manganese. Human PrimPol was incubated for 30 min with DNA substrate and each of the dNTPs in the presence of manganese. (**A**) Wild-type PrimPol was unable to extend from a ds DNA template with a blunt end, but could extend from a primer annealed to an overhanging template, even synthesising long tracts of homopolymers. (**B**) PrimPol$_{ZF-KO}$ could extend from an overhanging DNA template in the presence of manganese and, consistent with the wild-type, did not extend from a ds DNA substrate. The incorporation of 1 or 2 nucleotides of guanine or adenine opposite an overhanging template suggests that PrimPol$_{ZF-KO}$ incorporates in a low-fidelity template-dependent manner when incubated with overhanging DNA. (**C**) PrimPol$_{1–487}$ exhibited a highly similar terminal transferase activity spectrum to the PrimPol$_{ZF-KO}$. In the presence of manganese, it incorporated bases opposite an overhang in a low fidelity, template-dependent manner. (**D**) PrimPol$_{1–354}$ also exhibited low fidelity extension of a primer annealed to an overhanging template in the presence of manganese.

When this experimental strategy was repeated to study potential promiscuous DNA polymerase activities in the PrimPol$_{ZF-KO}$ and PrimPol$_{1–487}$ variants, we observed minimal mis-incorporation of all four tested single nucleotides. Interestingly, when a catalytically active fragment (PrimPol$_{1–354}$) was tested, we observed more efficient mis-incorporation compared to PrimPol$_{ZF-KO}$ and PrimPol$_{1–487}$ variants. However, PrimPol$_{1–354}$ fragment was

Table 1. Extension, TLS and fidelity activities associated with full-length and mutated forms of human PrimPol

A

Activity	PrimPol	$PrimPol_{ZF\text{-}KO}$	$PrimPol_{1\text{-}487}$	$PrimPol_{1\text{-}354}$	$PrimPol_{372\text{-}560}$
Primase	✓	×	✓	×	×
Polymerase	✓	✓	✓	✓	×
Template-independent primer extension (Mn^{2+})	✓	×	×	×	×

B

Lesions	Wild-type PrimPol		$PrimPol_{1\text{-}354}$	
	Read-through	Base incorporation	Read-through	Base incorporation
CPD	×	-	✓	A → A
6-4 Photoproduct	✓	T → C/G	✓	A → C/G C → A/C/G
8-oxoguanine	✓	A/C	✓	A/C
AP Site	×	-	×	
Deoxyuracil	✓	A	✓	A
Thymine Glycol	×	-	×	

Summary of the DNA synthesis activities of wild-type and mutant PrimPol. (**A**) PrimPol exhibits primase activity, polymerase activity and, in the presence of manganese, template-independent primer extension. Whilst the polymerase activity and template-independent primer extension activity appear to reside in the AEP polymerase domain alone, the primase activity requires the associated zinc finger. (**B**) PrimPol can read through a number of common DNA lesions and we have determined some of the lesions PrimPol can read through and which bases are incorporated opposite these lesions. As the UV lesions span two adjacent bases, a '→' is used to separate the bases incorporated opposite the first and second damaged bases.

able to mis-incorporate between 1 and 2 incorrect incoming nucleotides. We are unable to conclude if this unique enzymatic activity of human PrimPol is significant for the biological functions of this enzyme *in vivo*. We speculate that, under the experimental conditions tested, the Zfn module allowed manganese-dependent primer extension resulting in the synthesis of DNA that is non-complementary to the template strand. This hypothetical scenario is only possible when the enzyme was able to bind to a single-stranded template strand of the DNA primer-template substrate. This interesting observation has raised questions regarding the mechanism by which PrimPol binds to and coordinates the primer-template substrate during DNA polymerisation.

DNA binding activities of human PrimPol

Next, we analysed the binding of the AEP polymerase and Zfn domains of PrimPol to a number of DNA substrates to determine whether they bind specifically to ss or ds DNA. The Zfn domain is presumed to be a DNA-recognition domain, in addition to contributing to its primase activity, as previous studies demonstrated altered DNA binding activity associated with mutations in a related UL52 Zfn domain (8,20). EMSAs were performed to analyse the binding capabilities of the AEP polymerase ($PrimPol_{1–354}$) and Zfn ($PrimPol_{372–560}$) domains. $PrimPol_{1–354}$ showed a similar binding capacity to ds and ss DNA, suggesting that the polymerase can bind both substrates (Figure 5A). The observed DNA binding specificity reflects the likely presence of two closely clustered binding sites allowing recognition of a primer-template junction, which is an essential requirement for all DNA polymerases. Observed binding values for both ss DNA and ds DNA were evident at a minimum of ∼500 nM DNA and a complete shift of the DNA was evident at 5 μM DNA. This relatively weak coordination of DNA substrates is consistent with previously postulated biological functions of human PrimPol (4–7).

Analogous EMSA experiments were also performed to analyse the DNA binding capacity of the Zfn domain ($PrimPol_{372–560}$). In contrast with the AEP domain, PrimPol's Zfn domain was unable to bind ds DNA but did show specific binding to ss DNA (Figure 5B). This intriguing result provides experimental evidence supporting the potential importance of the Zfn module in recognition/coordination of the single-stranded template DNA downstream of the primer-template junction. The significance of this interaction will be discussed below. To determine if DNA binding was performed by the Zfn, we assayed a Zfn knockout domain ($PrimPol_{372–560/ZF\text{-}KO}$) for ss DNA binding activity (Figure 5C) and observed an absence of binding with this mutant, confirming the importance of the Zfn motif for nucleic acid recognition.

PrimPol's AEP domain possesses TLS activities

PrimPol has the capacity to bypass a number of replication stalling lesions, including UV ((6–4)PP) (4) and oxidative lesions (4,6). To address whether the TLS activities of PrimPol are independent of the zinc finger domain, we assayed the TLS activities of $PrimPol_{1–354}$ on a range of templates containing a variety of DNA modifications (Table 1B). As full-length PrimPol appears unable to read through CPDs *in vitro*, we assayed $PrimPol_{1–354}$ for TLS bypass of this common photolesion. Notably, although $PrimPol_{1–354}$ slowed

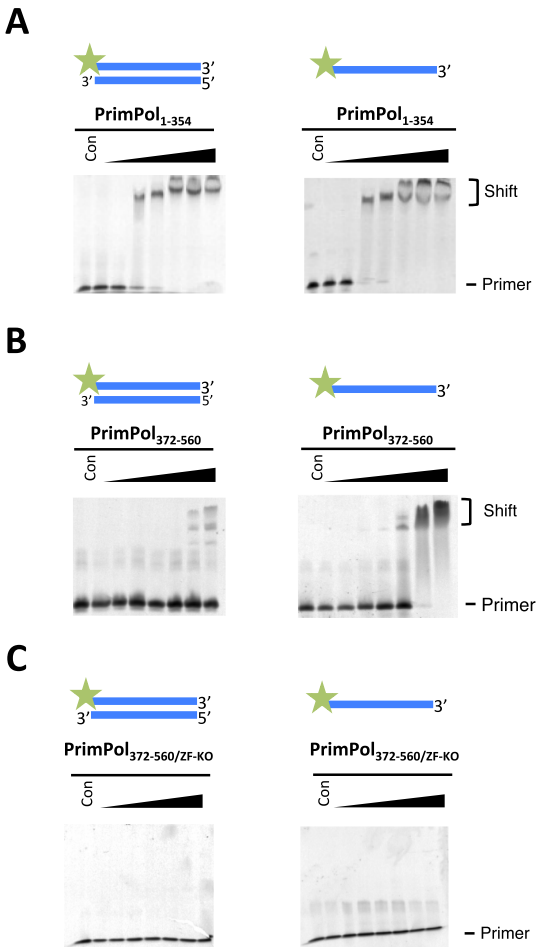


Figure 5. PrimPol's polymerase domain binds ss and ds DNA but the zinc finger domain binds only ss DNA. (**A**) PrimPol$_{1–354}$ was tested for its ability to bind DNA by shift assays. PrimPol polymerase domain was incubated with 40 nM DNA at various protein concentrations (0.05, 0.1, 0.5, 1.0, 5.0, 10.0 and 20.0 μM) for 60 min at 25°C. The polymerase domain binds to ds DNA and ss DNA with approximately equal proficiency. Binding was evident at concentrations ∼0.5 μM but a complete shift was observed at 5.0 μM for each DNA substrate tested. (**B**) PrimPol$_{372–560}$ (Zfn domain) was also assayed for binding to DNA activity. It was incubated with 40 nM DNA at various protein concentrations (0.05, 0.1, 0.5, 1.0, 5.0, 10.0 and 20.0 μM). A DNA shift was not observed with ds DNA substrate but a clear shift was evident with ss DNA, suggesting that the zinc finger binds single-stranded regions of DNA. (**C**) PrimPol$_{372–560/ZF-KO}$, lacking a functional Zfn, was incubated with 40 nM DNA at the same protein concentrations and ss DNA binding activity was not observed.

as it reached a CPD lesion, it showed the capacity to read through this bulky lesion (Figure 6A). In contrast, wild-type PrimPol cannot perform TLS opposite this lesion (4,5). Single nucleotide incorporation assays were employed to determine which bases are incorporated opposite CPDs, revealing that PrimPol performs error-free bypass of this photolesion (Figure 6A). A similar TLS activity was also confirmed with *Xenopus* $PrimPol_{1–334}$, which was also capable of reading through CPDs in an error-free manner (Supplementary Figure S1A). This TLS activity is similar to Pol η, a specialised TLS polymerase known to bypass CPDs in an error-free manner (21,22). As this read through of a CPD is only evident in the absence of the Zfn domain, it may be the case that PrimPol requires a conformational change or partner to assist in promoting this lesion bypass activity *in vivo*.

(6–4)PPs are a much more highly distorting UV lesion than CPDs. Pol η can incorporate bases opposite (6–4)PPs but incorporates an incorrect dG opposite the 3′ dT of the lesion and is unable to subsequently extend from this product (23). However, we previously reported that PrimPol exhibits the ability to incorporate opposite and extend through (6–4)PPs (4). Although PrimPol's TLS activity was retained when the Zfn was deleted, the nucleotide incorporation signature was altered (Figure 6B), again consistent with the proposed regulatory role of the Zfn domain. Wild-type PrimPol incorrectly incorporates thymine opposite the first dT of the (6–4)PP, whereas $PrimPol_{1–354}$ incorporates dA or dC opposite this base. If dA nucleotide is incorporated opposite the first thymine of the (6–4)PP, $PrimPol_{1–354}$ then incorporates either dC or dT. If the first nucleotide added is dC, it will then incorporate dA, dC or dT. To confirm whether this altered fidelity of nucleotide incorporation was orthologue-specific, we assayed the equivalent *Xenopus* PrimPol domain ($XPrimPol_{1–334}$) and observed showed a similar fidelity to the human PrimPol polymerase domain (Supplementary Figure S1B).

Next, we examined TLS bypass of an oxidative lesion, 8-oxoG. In common with wild-type enzyme, $PrimPol_{1–354}$ can read through 8-oxoG lesions with minimal stalling and exhibited the ability to insert A and C opposite 8-oxoG lesions (Figure 7A), as previously observed for wild-type PrimPol (4). A similar TLS activity and fidelity for this lesion was also evident with $XPrimPol_{1–334}$ (Supplementary Figure S2A). Cytosine in DNA is susceptible to spontaneous deamination to deoxyuracil (dU) *in vivo*, which can base pair to adenine, resulting in C→T transition mutations (10,11,24). Deoxyuracil is not recognised by, nor does it stall, eukaryotic TLS polymerases (10,13,14,25). Similarly, the polymerase domain of PrimPol does not stall when confronted with a dU base (Figure 7B). PrimPol incorporates an adenine opposite dU therefore, as one might expect, it reads the dU base as a thymine. This is in common with $XPrimPol_{1–334}$ (Supplementary Figure S2B) and wild-type PrimPol (Supplementary Figure S3).

Finally, we examined TLS bypass of apurinic/apyrimidinic (AP sites) and thymine glycol (TG) lesions. $PrimPol_{1–354}$ was unable to extend through either AP (Supplementary Figure S4A) or TG (Supplementary Figure S4B) lesions but there was some single nucleotide incorporated opposite the TG lesion. This insertion activity could potentially promote extension by another polymerase that is capable of extending following nucleotide insertion opposite the TG. $XPrimPol_{1–334}$ was also unable to read through either of these lesions (Supplementary Figure S4). The inability of $PrimPol_{1–354}$ to read through AP sites and TG lesions is also evident in wild-type PrimPol (4). This inability to extend through AP and TG sites, as well as other lesions, does not preclude the possibility that PrimPol reprimes following these lesions, allowing replicative polymerases to resume synthesis in a post-lesion fashion, see below.

Human PrimPol is a distributive DNA polymerase

We next analysed the number of nucleotides incorporated by human PrimPol in a single association-dissociation event to measure the processivity of this polymerase. In order to determine the processivity of PrimPol, we performed primer extension assays in the presence of an excess of DNA (herring sperm), which acted as a 'substrate trap' for polymerases that dissociate from the primer-template substrate. Using this approach, we identified that PrimPol inserts up to 4 nucleotides opposite an undamaged template, rarely inserting more than this (Figure 8A). When the Zfn domain was removed from PrimPol, the maximum number of nucleotides incorporated remained at ∼4 but the average number of nucleotides incorporated was higher, with fewer polymerases inserting 1–3 nucleotides (Figure 8B). We also carried out processivity assays in the presence of Taq polymerase to ensure that it was not the trap DNA that kept the polymerase from inserting more than 4 nucleotides. We observed that Taq polymerase possessed high processivity and incorporated nucleotides to the end of the template DNA (Supplementary Figure S5). Therefore, the Zfn appears to play a role in regulating the number of nucleotides that PrimPol inserts, possibly by stabilizing the interaction of the polymerase domain with the template, see the discussion.

PrimPol's zinc finger domain is required for bypass of lesions *in vivo*

We previously reported that avian *PrimPol*$^{-/-}$ cells showed significantly decreased rates of replication fork progression (4). To evaluate the requirement for the Zfn domain of human PrimPol *in vivo*, we produced cell lines, in a DT40 *PrimPol*$^{-/-}$ knockout background (4), stably expressing human PrimPol, $PrimPol_{ZF\text{-}KO}$, $PrimPol_{AXA}$ (catalytically null) or $PrimPol_{1–354}$ (Figure 9A). Using DNA fibre analysis, we confirmed that normal replication fork rates was restored by adding back human PrimPol, however this replication fork defect was not corrected by adding back the catalytically dead $PrimPol_{AXA}$ mutant (Figure 9B). Next, we studied the replication fork speed in $PrimPol_{1–354}$ and $PrimPol_{ZF\text{-}KO}$ complemented clones and observed that both proteins were able to restore replication fork rates to near wild-type levels (Figure 9B). This suggests that the polymerase activity of PrimPol is largely responsible for the maintenance of fork progression in unperturbed replication. Indeed, $PrimPol_{1–354}$ complemented cells showed even greater average replication fork rates compared to wild-type

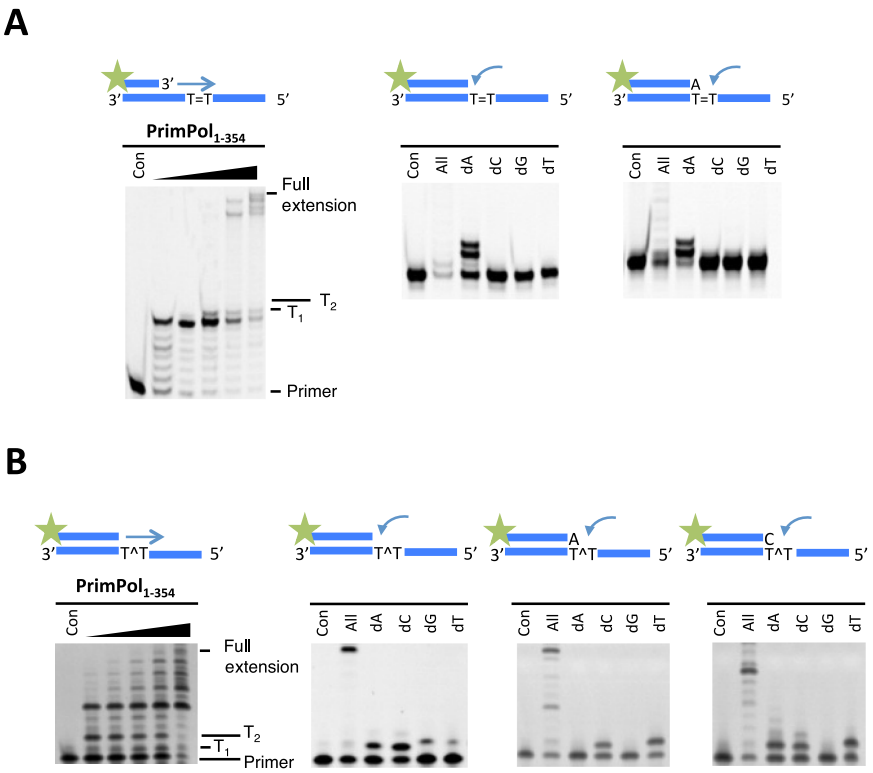


Figure 6. PrimPol$_{1–354}$ can replicate through CPD and (6–4(PP)) lesions. (**A**) PrimPol$_{1–354}$ was incubated with a primer-template substrate in which the template contained a CPD lesion downstream of the primer-template junction in the presence of dNTPs. Time points were taken between 0.5 and 60 min. PrimPol$_{1–354}$ extends from the primer up to the CPD, before stalling, it will then added a base opposite the first thymine of the CPD and continue to extend until the end of the template (left panel). PrimPol$_{1–354}$ was then incubated with each dNTP to test which nucleotide it incorporates opposite a CPD (right panels). PrimPol$_{1–354}$ incorporated adenine opposite the first and second thymine of the CPD. (**B**) The polymerase domain of PrimPol can also perform TLS bypass of a 6–4(PP) lesion immediately downstream of the primer-template junction (left panel). PrimPol$_{1–354}$ incorporates either an adenine or cytosine nucleotide opposite the first thymine of the 6–4(PP) (right panels). If an adenine is incorporated opposite the first thymine, a cytosine or thymine is then incorporated opposite the second. If a cytosine is incorporated opposite the first thymine, an adenine, cytosine or thymine will be incorporated opposite the second thymine of the 6–4(PP).

cells (Figure 9B), fitting with the superior polymerase activity observed *in vitro* (Figure 3A).

To examine the ability of the different PrimPol variants to support bypass of bulky lesions during replication *in vivo*, a large pulse of UV irradiation (20 J/m^2) was introduced to cells between the two labelling steps to investigate fork stalling induced at UV lesions. PrimPol knockout cells show a pronounced stalling phenotype, indicated by a large increase in the ratio of the CldU to IdU ratios (4). Consistent with these previous findings, we observed complementation by adding back the wild-type enzyme. However, PrimPol$_{AXA}$, PrimPol$_{1–354}$ or PrimPol$_{ZF\text{-}KO}$ were unable to correct this fork-stalling defect (Figure 9C). Therefore, although the polymerase extension activity of PrimPol is sufficient for maintenance of unperturbed replication, these data suggest that its primase activity is requisite for timely replication fork progression following the introduction of significant DNA damage.

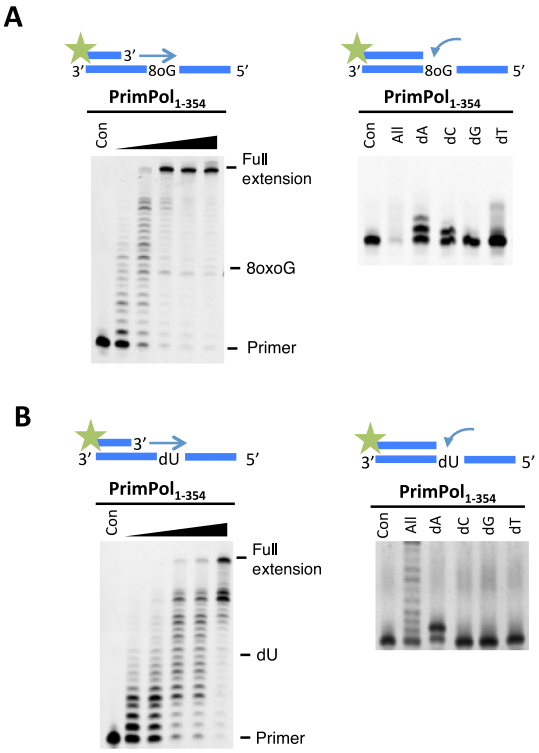


Figure 7. PrimPol$_{1–354}$ can replicate through an 8-oxoguanine (8oxoG) lesion and a deoxyuracil (dU) base. (**A**) The polymerase domain exhibited minimal stalling opposite an 8oxoG lesion and extended fully to the end of the template. PrimPol incorporated an adenine or cytosine opposite the 8oxoG, which are the expected bases to be incorporated by Hoogsteen base pairing or Watson–Crick base pairing, respectively. (**B**) PrimPol$_{1–354}$ efficiently synthesised through a templating dU site. PrimPol incorporates an adenine nucleotide opposite the dU lesion through Watson–Crick base pairing.

DISCUSSION

Recent studies have identified that PrimPol is a novel eukaryotic DNA polymerase capable of performing both priming and primer extension synthesis (4–7). It is likely that PrimPol was acquired early in eukaryotic evolution to play a significant role in lesion bypass during DNA replication. PrimPol-mediated damage tolerance appears to provide additional mechanisms that facilitate replication fork progression following stalling of replisomes at sites of DNA damage or secondary structure.

In this study, we have dissected the molecular activities associated with the major structural elements of human PrimPol (Table 1A) and determined the cellular requirements for these activities *in vivo*. PrimPol is both an active primase and polymerase with the capacity to perform TLS (Table 1B). An intact zinc finger is essential for PrimPol's primase activity *in vitro* (Figure 2), consistent with reports for a similar requirement for many diverse DNA primases (8,9,26,27). It has been proposed that this structural element is required for template recognition and regulation of primer length synthesis (28). Although the DNA polymerase activity of PrimPol appeared initially to be independent of the zinc finger domain, as a catalytically active domain lacking the zinc finger domain (PrimPol$_{1–354}$) showed enhanced relative polymerase and processivity activities compared to the wild-type enzyme. Further studies revealed that the zinc finger modulates the extension activities and fidelity of PrimPol and therefore it should also

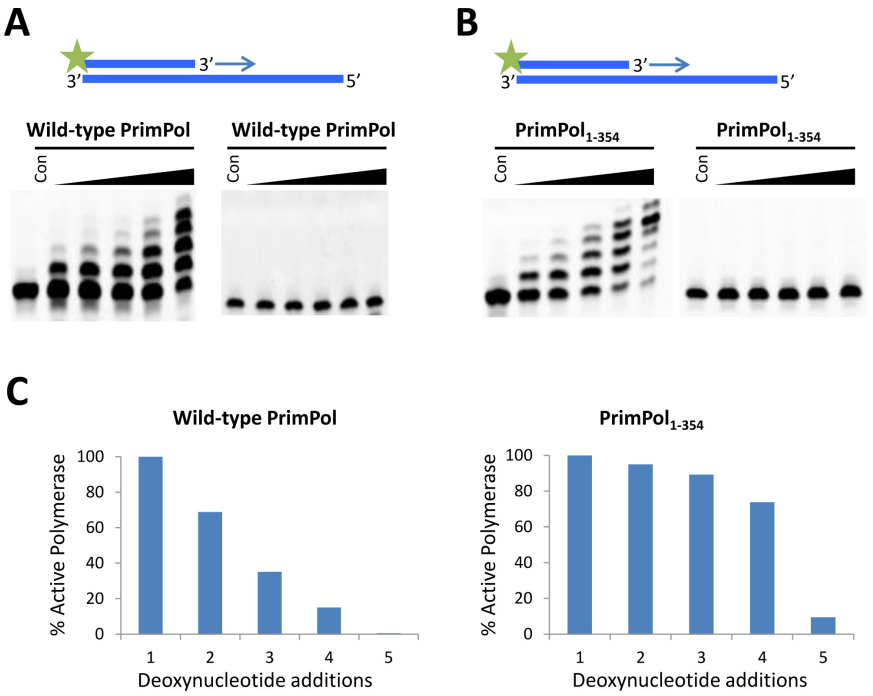


Figure 8. Processivity of the polymerase activity of PrimPol. PrimPol was pre-incubated for 30 min at 37°C with an undamaged DNA primer-template substrate to allow PrimPol to bind to the DNA. The reaction was initiated by the addition of dNTPs and an excess of sonicated herring sperm DNA (trap) and time points taken at 15, 30, 60, 120 and 360 s. (**A**) After 360 s, wild-type PrimPol incorporated up to 4 nucleotides opposite the template but a significant fraction of enzyme incorporated only 1, 2 or 3 nucleotides (left panel). To confirm that the trap prevents polymerase extending from a second template, the trap was also added into the pre-incubation mix with PrimPol and the DNA substrate. This reaction was supplemented with dNTPs and there is no extension (right panel), thus successfully exhibiting the effectiveness of the trap. (**B**) $PrimPol_{1–354}$ also predominantly incorporates up to 4 nucleotides but there were fewer polymerases incorporating only 1, 2 or 3 nucleotides (left panel). The effectiveness of the trap was also successfully confirmed (right panel). (**C**) Percentage of PrimPol molecules incorporating at least *n* dNTPs for either full-length PrimPol or $PrimPol_{1–354}$, calculated using Equation (1).

be considered to be an important regulator of PrimPol's polymerase functions. PrimPol is a very distributive enzyme, catalysing up to four nucleotide incorporations during a single DNA substrate binding/polymerisation cycle *in vitro*. This distributive character of PrimPol is reminiscent of the DNA polymerase processivity reported for the Y-family TLS polymerases, such as Pol κ and η, which predominantly incorporate up to 6 or 7 nucleotides (16,29).

Although removal of the last 80 amino acids at the C-terminal region of human PrimPol ($PrimPol_{1–487}$) did not abolish primase activity, which suggests that the C-terminus is not an absolute requirement for activity, it did reduce activity to some extent. Thermal denaturation studies have shown that there is a considerable increase in the stability of PrimPol upon removal of amino acids 488–560 (Supplementary Figure S6), suggesting that this region lacks discernible structure that would contribute to protein stability. Disordered regions of proteins often serve as scaffolds involved in protein–protein interactions and we speculate that this portion of human PrimPol might also act as an interaction interface. Indeed, it has been reported that the carboxyl-terminus of PrimPol mediates an interaction with the large subunit of human replication protein A (6).

In the presence of manganese, PrimPol appeared to be much more active than with magnesium. However, upon closer inspection, we discovered that the enzyme can produce polymeric primer extension products, even when only a single form of dNTP is present. Under these manganese-

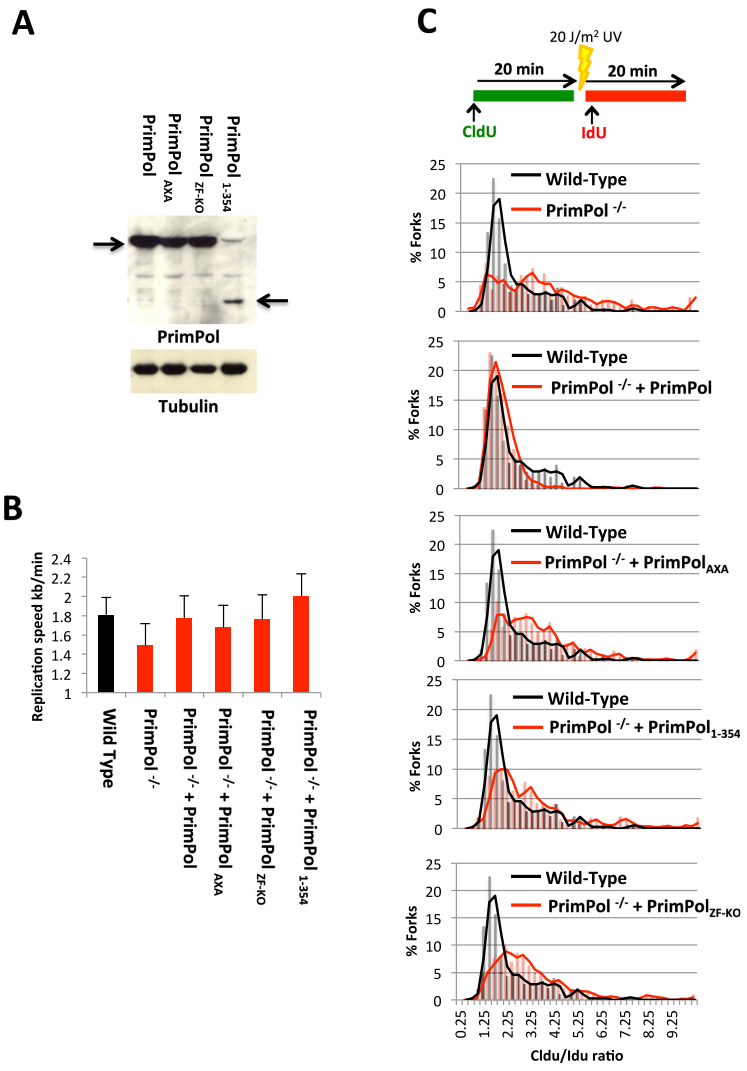


Figure 9. PrimPol's zinc finger domain is required for UV lesion bypass *in vivo*. (**A**) *PrimPol*$^{-/-}$ DT40 cells were complemented with variants of PrimPol and their *in vivo* expression confirmed by western blot. (**B**) Unperturbed replication fork rates were analysed by DNA fibre analysis using CldU and IdU labels for all the complemented cell types. The data represent the mean of three or more experiments with error bars showing the mean standard deviation. (**C**) DNA fork rate analysis was also carried out with after a 20 J/m^2 pulse of UV, between CldU and IdU labels, to look for fork-stalling. Data is shown as the ratio of the two labels, Cldu, pre-UV : IdU, post UV label ($n = 3$). The black lines indicate the Cldu:IdU ratio of the two labels in the wild-type cells and the red lines indicate the ratios in cells complemented with the indicated PrimPol variants.

dependent conditions, PrimPol extends from a primer to form homopolymeric strands that are non-complementary and therefore unlikely to be annealed to the template strand. This suggests that caution should be taken when interpreting the results from studies using manganese as the resulting products may represent aberrant synthesis products rather than the expected template-dependent ones. Manganese has the effect of decreasing fidelity amongst polymerases (30,31). Whilst magnesium enforces tetrahedral geometry in the arrangement of its ligands, manganese will accommodate square, planar, tetrahedral and octahedral coordinations thus increasing the ability to accelerate the rate of reaction in substrates that are perhaps misaligned (32). The physiological implications of this template-independent primer extension activity of PrimPol in the presence of manganese remain unclear, as the cellular concentrations of manganese are low, suggesting this activity may be irrelevant *in vivo* or it may be limited to certain cellular compartments (e.g. mitochondria).

The translesion bypass activities associated with wild-type PrimPol and $PrimPol_{1–354}$ are compared in Table 1B. As PrimPol can incorporate bases opposite and extend from highly distorting lesions, it is likely that the active site involved in production of phosphodiester bond is large and flexible to accommodate a range of lesions. The (6–4) photoproducts, in particular, are highly mutagenic due to the high degree of distortion they introduce into the DNA backbone. The helical bending at these lesions is estimated to be ∼44° and the distance between phosphorus atoms in the paired bases in the (6–4)PP is ∼6 Å, compared to the 7 Å average corresponding distance at other base-pairing steps (33). Therefore, it is probably unsurprising that PrimPol exhibits low fidelity when inserting bases opposite a (6–4)PP, as a polymerase that can accommodate this lesion would require a flexible active site. Unexpectedly, the polymerase domain of PrimPol, like Pol η, is capable of bypassing CPDs with high fidelity. This contrasts with full-length PrimPol, which is not capable of bypassing CPDs unless two adenines are inserted opposite the CPD (4,5). This result establishes that the active site of PrimPol has the capacity to bypass these lesions but this is constrained in the full-length enzyme. A conformation change may be required, potentially induced by binding partners, such as Replication Protein A (RPA), that alters the activity of PrimPol to allow it to bind to and bypass CPDs *in vivo*.

PrimPol is also confronted with a number of lesions as a result of oxidative stress in cells. These include AP sites and 8-oxoG lesions. Pol γ, the mitochondrial replicative polymerase, has the ability to bypass an 8-oxoG site but at a reduced rate, and cannot bypass AP sites (34,35). Conversely, PrimPol shows no hindrance to incorporate bases opposite 8-oxoG lesions and, although it cannot incorporate bases opposite an AP site, it has the capacity to re-prime post lesion. 8-oxoG is expected to pair with C through Watson–Crick base pairing in its regular *anti* conformation, but in its *syn* conformation it forms Hoogsteen base pairs with adenine (36). This ability to insert A, as well as C, opposite 8-oxoG is apparent in a number of polymerases but the addition of auxiliary proteins (e.g. RPA and PCNA) can dramatically increase the preference for adenine over cytosine, as is the case with Pol λ and η (36,37). It remains to be seen whether this is also the case with PrimPol. This ability to accommodate bulky lesions comes at the expense of fidelity and, therefore, PrimPol is a low fidelity polymerase that is likely to make limited, and perhaps, nonspecific contacts with the replicating base pair. Deoxyuracil is not recognised as a lesion by replicative polymerases and results in C:G → T:A transition mutations. Deoxyuracil is recognised by archaeal family B polymerases as a lesion and stalls replication to prevent mutation (38) but PrimPol does not have this recognition capacity and incorporates an adenine opposite a deoxyuracil lesion.

Canonical DNA polymerases share a common right-handed architecture containing fingers, palm and thumb subdomains. All of the three structural elements are clustered together providing accurate, highly structurally restrained positioning of the polymerase at the primer-template junction in order to promote the highest fidelity and processivity of DNA synthesis. Although this tight and specific binding cleft is perfectly suited for replication of long stretches of DNA, in some instances such structural arrangements can be disadvantageous. In particular, it limits the capacity of the replicative enzymes to tolerate abnormalities in the topology of DNA templates, such as secondary structures or DNA damage-induced distortions.The thumb subdomain of replicative polymerases interacts with the sugar-phosphate backbone of the DNA to coordinate the DNA. This subdomain is optimised for the highly accurate replication of DNA and is sensitive for structural anomalies in the in the DNA template. The predicted structure of PrimPol, in addition to the known structures of primases, suggests that this polymerase lacks structural elements corresponding to the 'thumb' domain that is present in the replicative polymerases that is implicated in processivity (39). The open planar predicted structure of this polymerase indicates a transient binding of the polymerase domain and this is consistent with other low fidelity polymerases, such as those of the Y-family of polymerases (40). The processivity data presented here suggests that the zinc finger domain of PrimPol regulates the fidelity of these polymerases. Whilst there is no reduction in the activity of the polymerase domain lacking the zinc finger, it actually appears more active, we propose that the presence of the zinc finger reduces the processivity of the enzyme to allow a slower, higher fidelity incorporation of complementary nucleotides. We postulate that whilst the AEP polymerase domain provides a platform to recognise the primer-template junction of the DNA substrate, the zinc finger binds the single-stranded region of the template strand to modulate binding and fidelity by the AEP domain (Figure 10A). Y-family polymerases have a highly truncated thumb domain and, notably, contain an additional functional module referred to as the polymerase associated domain (PAD). Interestingly, studies on TLS polymerases Dpo4 and Dbh point to a role for PAD in regulation of processivity and fidelity of these enzymes (41). This structural element not only binds ss DNA and provides an extended flexible platform coordinating the template strand during DNA synthesis on damaged DNA templates, but also modulates the activity of these TLS DNA polymerases. As part of an adaptation to perform TLS, we suggest that the zinc finger is functionally

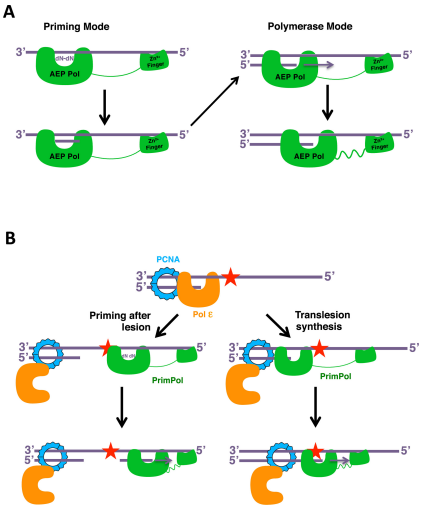


Figure 10. Models of the mechanisms of action of PrimPol. (**A**) PrimPol has both DNA primase and polymerase activities. The zinc finger of PrimPol stabilises the binding of the polymerase domain on single-stranded DNA. The AEP polymerase domain coordinates two adjacent deoxynucleotides opposite the single-stranded DNA template and the stability afforded to this complex by the Zfn binding allows synthesis of a dinucleotide. This dinucleotide can then be extended in a more processive way to form longer products. PrimPol can also bind primer-template junctions to catalyse more canonical primer extension reactions. Both the polymerase and primase activities are intrinsic to the AEP domain. Although the zinc finger is essential for maintaining less 'stable' primer synthesis, it is also important for modulating the fidelity and processivity of the polymerase activities. This stabilizing influence of the Zfn may be particularly important for allowing productive synthesis during TLS bypass of distorting lesions and structures. We postulate that the primase and polymerase activities of PrimPol are not discrete activities but rather represent different synthesis modes performed under different DNA binding conditions. (**B**) PrimPol uses these activities to bypass DNA damage that blocks replicative polymerases. When replicative polymerases encounter a blocking lesion, they are displaced or regress from the site of damage. PrimPol can then access the template upstream of the lesion, where it can reprime following the lesion (left) or if it can access the primer-template junction it can carry out TLS (right). Once PrimPol-mediated bypass of the lesion has occurred, the replicative polymerase can then resume replication at the fork.

analogous to the PAD subdomain, but is spatially separated from the polymerase domain to allow greater freedom of binding of the enzyme to distorted DNA templates, such as those containing UV-induced lesions.

The activity most closely associated with the primase superfamily, namely primer synthesis, is considered to be somehow different from primer extension activities associated with canonical polymerases. However, primases should be considered to be polymerases that have evolved to catalyse a poorly processive extension reaction involving the synthesis of a dinucleotide. To do this, they have acquired the capacity to bind a single nucleotide as the 'primer', which provides the 3′OH to initiate attack on a second template-bound nucleotide to form the first phosphodiester linkage. As Zfn binds single-stranded DNA and its absence negates primer synthesis, we propose that the Zfn acts in concert with the polymerase domain to stabilise the binding and configuration of the template, polymerase and nucleotides long enough to allow the production of the initial dinucleotide 'primer' (Figure 10A). This product is subsequently extended in a more processive polymerisation mode. In replicative primases, the primase large subunit probably provides a similar role (12). In addition to primer synthesis, the zinc finger domain of PrimPol also plays additional roles in regulating the fidelity of these enzymes. By lowering the processivity of PrimPol, it appears to increase its fidelity, thus reducing its potential to be pro-mutagenic.

Finally, regarding the bypass of lesions during genome replication *in vivo*, this and previous work (4) suggest that both of PrimPol's priming and extension activities are requisite for lesion bypass and replication restart. When replicative DNA polymerases encounter a blocking lesion, they stall and PrimPol is recruited, possibly by RPA, to perform TLS or re-prime replication downstream of the lesion (Figure 10B). After production of a primer or polymerisation through the lesion, the replicative polymerase can then resume replication. PrimPol's extension activities appear to be PCNA-independent as no stimulation was observable in the presence of this protein (unpublished data). Although it has been reported that RPA can bind to and potentially recruit PrimPol to DNA, the mechanism by which it interfaces with stalled replisomes to restart fork progression remains to be established.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to kindly thank Chris Dadswell and Sarah Parry-Morris for their technical assistance. We would also like to thank Prof. Shigenori Iwai who kindly provided DNA containing a 6–4 photoproduct.

FUNDING

Biotechnology and Biological Sciences Research Council (BBSRC) [BB/H019723/1]; MRC [G080130]; BBSRC PhD studentship (BB/F01645X/1 to B.A.K.). Source of open access funding: RCUK.

Conflict of interest statement. None declared.

REFERENCES

1. Frick,D.N. and Richardson,C.C. (2001) DNA primases. *Annu. Rev. Biochem.*, **70**, 39–80.
2. Aravind,L., Leipe,D.D. and Koonin,E.V. (1998) Toprim-a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.*, **26**, 4205–4213.
3. Iyer,L.M., Koonin,E.V., Leipe,D.D. and Aravind,L. (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.*, **33**, 3875–3896.

4. Bianchi,J., Rudd,S.G., Jozwiakowski,S.K., Bailey,L.J., Soura,V., Taylor,E., Stevanovic,I., Green,A.J., Stacker,T.H., Lindsay,H.D. *et al.* (2013) Eukaryotic PrimPol bypasses UV photoproducts during chromosomal DNA replication. *Mol. Cell*, **52**, 566–573.
5. Rudd,S.G., Glover,L., Jozwiakowski,S.K., Horn,D. and Doherty,A.J. (2013) PPL2 translesion polymerase is essential for the completion of chromosomal DNA replication in the African Trypanosome. *Mol. Cell*, **52**, 554–565.
6. Wan,L., Lou,J., Xia,Y., Su,B., Liu,T., Cui,J., Sun,Y., Lou,H. and Huang,J. (2013) hPrimpol1/CCDC111 is a human DNA primase-polymerase required for the maintenance of genome integrity. *EMBO Rep.*, **14**, 1104–1112.
7. Garcia-Gomez,S., Reyes,A., Martínez-Jiménez,M.I., Chocrón,E.S., Mourón,S., Terrados,G., Powell,C., Salido,E., Mendez,J., Holt,I.J. *et al.* (2013) PrimPol, an Archaic Primase/Polymerase Operating in Human Cells. *Mol. Cell*, **52**, 541–553.
8. Biswas,N. and Weller,S.K. (1999) A mutation in the C-terminal putative Zn2+ finger motif of UL52 severely affects the biochemical activities of the HSV-1 helicase-primase subcomplex. *J. Biol. Chem.*, **274**, 8068–8076.
9. Chen,Y., Carrington-Lawrence,S.D., Bai,P. and Weller,S.K. (2005) Mutations in the putative zinc-binding motif of UL52 demonstrate a complex interdependence between the UL5 and UL52 subunits of the human herpes simplex virus type 1 helicase/primase complex. *J. Virol.*, **79**, 9088–9096.
10. Lao-Sirieix,S.-H. and Bell,S.D. (2004) The heterodimeric primase of the hyperthermophilic archaeon Sulfolobus solfataricus possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. *J. Mol. Biol.*, **344**, 1251–1263.
11. Lipps,G., Weinzierl,A.O., Scheven,von, G., Buchen,C. and Cramer,P. (2004) Structure of a bifunctional DNA primase-polymerase. *Nat. Struct. Mol. Biol.*, **11**, 157–162.
12. Lao-Sirieix,S.-H., Nookala,R.K., Roversi,P., Bell,S.D. and Pellegrini,L. (2005) Structure of the heterodimeric core primase. *Nat. Struct. Mol. Biol.*, **12**, 1137–1144.
13. Bocquier,A.A., Liu,L., Cann,I.K.O., Komori,K., Kohda,D. and Ishino,Y. (2001) Archaeal primase: bridging the gap between RNA and DNA polymerases. *Curr. Biol.*, **11**, 452–456.
14. Liu,L., Komori,K., Sonoko,I., Bocquier,A.A., Cann,I.K.O., Kohda,D. and Ishino,Y. (2001) The Archaeal DNA Primase. Biochemical Characterization of the P41-P46 Complex From Pyroccocus Furiosis. *J. Biol. Chem.*, **276**, 45484–45490.
15. Engelke,D.R., Krikos,A., Bruck,M.E. and Ginsburg,D. (1990) Purification of Thermus aquaticus DNA polymerase expressed in Escherichia coli. *Anal. Biochem.*, **191**, 396–400.
16. Washington,M.T., Johnson,R.E., Prakash,S. and Prakash,L. (1999) Fidelity and processivity of Saccharomyces cerevisiae DNA polymerase eta. *J. Biol. Chem.*, **274**, 36835–36838.
17. Shannon,R.D. (1976) Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A*, **32**, 751–767.
18. Della,M., Palmbos,P.L., Tseng,H.-M., Tonkin,L.M., Daley,J.M., Topper,L.M., Pitcher,R.S., Tomkinson,A.E., Wilson,T.E. and Doherty,A.J. (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science*, **306**, 683–685.
19. Pitcher,R.S., Tonkin,L.M., Green,A.J. and Doherty,A.J. (2005) Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. *J. Mol. Biol.*, **351**, 531–544.
20. Chen,Y., Livingston,C.M., Carrington-Lawrence,S.D., Bai,P. and Weller,S.K. (2007) A mutation in the human herpes simplex virus type 1 UL52 zinc finger motif results in defective primase activity but can recruit viral polymerase and support viral replication efficiently. *J. Virol.*, **81**, 8742–8751.
21. Masutani,C., Kusumoto,R., Yamada,A., Dohmae,N., Yokoi,M., Yuasa,M., Araki,M., Iwai,S., Takio,K. and Hanaoka,F. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature*, **399**, 700–704.
22. Biertümpfel,C., Zhao,Y., Kondo,Y., Ramón-Maiques,S., Gregory,M., Lee,J.Y., Masutani,C., Lehmann,A.R., Hanaoka,F. and Yang,W. (2010) Structure and mechanism of human DNA polymerase η. *Nature*, **465**, 1044–1048.
23. Johnson,R.E., Haracska,L., Prakash,S. and Prakash,L. (2001) Role of DNA polymerase η in the bypass of a (6–4) TT photoproduct. *Mol. Cell. Biol.*, **21**, 3558–3563.
24. Duncan,B.K. and Miller,J.H. (1980) Mutagenic deamination of cytosine residues in DNA. *Nature*, **287**, 560–561.
25. Connolly,B.A., Fogg,M.J., Shuttleworth,G. and Wilson,B.T. (2003) Uracil recognition by archaeal family B DNA polymerases. *Biochem. Soc. Trans.*, **31**, 699–702.
26. Mendelman,L.V., Beauchamp,B.B. and Richardson,C.C. (1994) Requirement for a zinc motif for template recognition by the bacteriophage T7 primase. *EMBO J.*, **13**, 3909.
27. Griep,M.A. and Lokey,E.R. (1996) The role of zinc and the reactivity of cysteines in Escherichia coli primase. *Biochemistry*, **35**, 8260–8267.
28. Corn,J.E., Pease,P.J., Hura,G.L. and Berger,J.M. (2005) Crosstalk between primase subunits can act to regulate primer synthesis in trans. *Mol. Cell*, **20**, 391–401.
29. Ohashi,E., Bebenek,K., Matsuda,T., Feaver,W.J., Gerlach,V.L., Friedberg,E.C., Ohmori,H. and Kunkel,T.A. (2000) Fidelity and processivity of DNA synthesis by DNA polymerase κ, the product of the human DINB1 gene. *J. Biol. Chem.*, **275**, 39678–39684.
30. Tabor,S. and Richardson,C.C. (1989) Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 4076–4080.
31. Hays,H. and Berdis,A.J. (2002) Manganese substantially alters the dynamics of translesion DNA synthesis. *Biochemistry*, **41**, 4771–4778.
32. Pelletier,H., Sawaya,M.R., Wolfle,W., Wilson,S.H. and Kraut,J. (1996) A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase beta. *Biochemistry*, **35**, 12762–12777.
33. Kim,J.-K. and Choi,B.-S. (1995) The solution structure of DNA duplex-decamer containing the (6-4) photoproduct of thymidylyl (3′→ 5′) thymidine by NMR and relaxation matrix refinement. *Eur. J. Biochem.*, **228**, 849–854.
34. Pinz,K.G., Shibutani,S. and Bogenhagen,D.F. (1995) Action of mitochondrial DNA polymerase gamma at sites of base loss or oxidative damage. *J. Biol. Chem.*, **270**, 9202–9206.
35. Graziewicz,M.A., Bienstock,R.J. and Copeland,W.C. (2007) The DNA polymerase gamma Y955C disease variant associated with PEO and parkinsonism mediates the incorporation and translesion synthesis opposite 7,8-dihydro-8-oxo-2'-deoxyguanosine. *Hum. Mol. Genet.*, **16**, 2729–2739.
36. van Loon,B., Markkanen,E. and Hübscher,U. (2010) Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. *DNA Repair*, **9**, 604–616.
37. Maga,G., Villani,G., Crespan,E., Wimmer,U., Ferrari,E., Bertocci,B. and Hübscher,U. (2007) 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. *Nature*, **447**, 606–608.
38. Fogg,M.J., Pearl,L.H. and Connolly,B.A. (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.*, **9**, 922–927.
39. Brautigam,C.A. and Steitz,T.A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.*, **8**, 54–63.
40. Lehmann,A.R., Niimi,A., Ogi,T., Brown,S., Sabbioneda,S., Wing,J.F., Kannouche,P.L. and Green,C.M. (2007) Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair*, **6**, 891–899.
41. Boudsocq,F., Kokoska,R.J., Plosky,B.S., Vaisman,A., Ling,H., Kunkel,T.A., Yang,W. and Woodgate,R. (2004) Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. *J. Biol. Chem.*, **279**, 32932–32940.

12102–12111 *Nucleic Acids Research, 2014, Vol. 42, No. 19* *Published online 27 September 2014*
doi: 10.1093/nar/gku879

Human PrimPol mutation associated with high myopia has a DNA replication defect

Benjamin A. Keen, Laura J. Bailey†, Stanislaw K. Jozwiakowski† and Aidan J. Doherty*

Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton BN1 9RQ, UK

Received June 20, 2014; Revised September 12, 2014; Accepted September 12, 2014

ABSTRACT

PrimPol is a primase-polymerase found in humans, and other eukaryotes, involved in bypassing lesions encountered during DNA replication. PrimPol employs both translesion synthesis and repriming mechanisms to facilitate lesion bypass by the replisome. PrimPol has been reported to be a potential susceptibility gene associated with the development of myopia. Mutation of tyrosine 89 to aspartic acid (PrimPol^{Y89D}) has been identified in a number of cases of high myopia, implicating it in the aetiology of this disorder. Here, we examined whether this mutation resulted in any changes in the molecular and cellular activities associated with human PrimPol. We show that PrimPol^{Y89D} has a striking decrease in primase and polymerase activities. The hydrophobic ring of tyrosine is important for retaining wild-type extension activity. We also demonstrate that the decreased activity of PrimPol^{Y89D} is associated with reduced affinities for DNA and nucleotides, resulting in diminished catalytic efficiency. Although the structure and stability of PrimPol^{Y89D} is altered, its fidelity remains unchanged. This mutation also reduces cell viability after DNA damage and significantly slows replication fork rates *in vivo*. Together, these findings establish that the major DNA replication defect associated with this PrimPol mutant is likely to contribute to the onset of high myopia.

INTRODUCTION

Cells are required to replicate their genomes in a faithful way to avoid mutagenesis thus maintaining genetic stability. However, DNA is highly prone to damaging agents, including oxygen-free radicals, ultraviolet (UV) light, ionizing radiation and desiccation, resulting in the production of lesions and breaks that disrupt the structure of the double helix. One of the many undesirable consequences of such damage is the disruption of the progression of DNA replicases (1), which copy the genetic material. These replication machines are exquisitely sensitive to the conformation of the template and are prone to stalling upon encountering DNA damage. To overcome such genetic obstacles, cells have evolved a variety of lesion tolerance pathways to allow bypass of damage and resumption of replication (2,3).

PrimPol is a recently characterized primase-polymerase, belonging to archaeo-eukaryotic primase (AEP) superfamily, that is involved in lesion bypass during both nuclear and mitochondrial replication in eukaryotic cells (4–6). PrimPol is required for the replication past particular lesions, such as UV photoproducts, that block the cellular replication machinery. It can deploy two different mechanisms, utilizing either of its polymerase and primase activities to facilitate efficient lesion bypass. PrimPol can replicate directly through a number of obstructing lesions by translesion DNA synthesis (TLS) and it can also reprime replication post-damage to facilitate fork restart (4,7,8), although the exact timing of their usage remains to be established. PrimPol knockout cells are sensitive to damaging agents (e.g. UV, 4NQO and MMS) and murine PrimPol$^{-/-}$ cells show significant levels of DNA breaks, particularly after treatment with agents that stall replication (4).

Although recent studies have established the importance of PrimPol in damage tolerance during DNA replication in cellular and animal model systems, its association with human diseases has only been conjectured. However, exome sequencing of the genomes of candidate patients has recently implicated the PrimPol gene as a possible susceptibility gene associated with high myopia (9). This study identified a missense mutation in human PrimPol caused by a thymine to guanine transversion, resulting in the substitution of tyrosine at position 89 by an aspartic acid residue (Y89D). This mutation was not only found in a single family but also four additional sporadic patients with high myopia and these patients had high refractive errors (−6.5 to −32.15 diopters, D) (9). Myopia, or short-sightedness, is a common problem around the world in those with visual defects and this problem is rapidly increasing. Myopia accounts for 41% of decreased vision in Western populations and this up from 25% in the early 1970s (10). In some South-East Asian populations up to 80–90% of all children leaving

*To whom correspondence should be addressed. Tel: +44 1273 877 500; Fax: +44 1273 678 121; Email: a.j.doherty@sussex.ac.uk
†The authors wish it to be known that, in their opinion, these two authors contributed equally to the work.

school are now myopic (11). Notably 10–20% of these children suffer from high myopia, an extreme form of the disease, often characterized by a refractive error of less than −6.00 D and is often associated with other serious ocular disorders (11,12). The aetiology of high myopia remains extremely complex and involves the synergy between a number of genes.

The active site of PrimPol is predicted to consist of three magnesium-coordinating, negatively charged residues: an aspartate at residue 114, a glutamate at residue 116 and an aspartate at residue 280 (4). Although the tyrosine at position 89 is not predicted to reside directly within the active site, it is in close proximity and therefore may play yet undetermined roles in the activities associated with these enzymes. The mutated tyrosine (Y89) is a highly conserved residue across a wide array of chordate species (Figure 1) and *PolyPhen-2* prediction programme indicates that mutation of this residue would be functionally damaging (9). In addition to PrimPol, this tyrosine is also present in the related predicted primase PF14_0050 in *plasmodium falciparum*, so its evolutionary conservation is indicative of a significant functional residue (13).

In this study, we aimed to characterize the biochemical properties of the Y89D mutant of PrimPol (PrimPol^{Y89D}) to determine if this amino acid change alters PrimPol's activities thus establishing a molecular basis for its proposed involvement in high myopia. Here, we demonstrate that the Y89D mutant loses its ability to prime using ribonucleotides (rNTPs), but can still prime using deoxyribonucleotides (dNTPs). PrimPol^{Y89D} has a greatly reduced template-dependent processivity activity, when compared with the wild-type enzyme. In addition, we report that this mutant has significantly reduced binding for both DNA and dNTPs but its fidelity remains largely unchanged. Furthermore, we show that it is the hydrophobic ring of the tyrosine at position 89 that is important for maintaining these activities. We also demonstrate that changes in enzyme activity are the result of global changes in the structure of the polymerase domain of PrimPol. Finally, we report that this mutation significantly slows replication fork rates *in vivo* and reduces cell viability after DNA damage. Together, these data establish that a point mutation identified in PrimPol from patients with high myopia results in a major disruption of the catalytic and replication activities associated with human PrimPol thus establishing a link between replication stress and human disease, particularly high myopia.

MATERIALS AND METHODS

Construction of human PrimPol mutants

Wild-type human PrimPol was cloned as described previously (4). PrimPol^{Y89D}, PrimPol^{Y89F} and PrimPol^{Y89S} were cloned by site-directed mutagenesis using the primers that can be found in Supplementary Table S1. PrimPol$_{1–354}$ was cloned as described previously (14) and PrimPol$^{Y89D}{}_{1–354}$ was cloned by site-directed mutagenesis from this construct using the primers in Supplementary Table S1.

Expression and purification of recombinant PrimPol proteins

Wild-type PrimPol, PrimPol^{Y89D}, PrimPol^{Y89F} and PrimPol^{Y89S} were expressed in BL21 (pLysS) *Escherichia coli* cells overnight at 16°C and PrimPol$_{1–354}$ and PrimPol$^{Y89D}{}_{1–354}$ were expressed overnight at 25°C. The proteins were then purified as described previously (14). Protein concentrations were determined using the absorbance at 280 nm and the extinction coefficient for each of the protein constructs.

DNA primase assays

DNA primase assays were carried out as described previously (14). The templating oligonucleotide sequences can be found as sequences 1–4 in Supplementary Table S2. Products were resolved on a 15% (v/v) polyacrylamide gel containing 7 M urea and 1× TBE buffer at 850 V for 2.5 h in 1× TBE buffer. Fluorescently labelled DNA oligomers were detected by scanning using a Fujifilm FLA-5100 image reader.

DNA primer extension assays

Polymerase activity was determined by DNA primer extension assays. Hex-labelled fluorescent primers were annealed to unlabelled templates (sequences in Supplementary Table S2). Note that 100 nM of enzyme was incubated at 37°C with 20 nM DNA, 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$, 1 mM DTT and 200 μM dNTPs (Roche) to a final volume of 20 μl for time points increasing up to 60 min. Reactions were quenched using 2× stop buffer (95% formamide, 0.09% xylene cyanol, 0.05% bromophenol blue, 200 nM competitor oligonucleotide) and boiled at 95°C for 5 min. Products were resolved by electrophoresis as in the DNA primase assays. For assessment of the fidelity of PrimPol$_{Y89D}$, oligonucleotides with two adjacent, identical bases downstream of the primer were incubated, along with 100 nM enzyme, with single nucleotides for 20 min at 37°C and were subsequently resolved as above.

Electrophoretic mobility shift assays (EMSAs)

EMSAs were carried out on an overhanging DNA substrate (sequences 5 and 6 annealed from Supplementary Table S2). Varying concentrations of PrimPol$_{1–354}$ and PrimPol$^{Y89D}{}_{1–354}$ were added to 60 nM DNA, 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$ and 1.0 mM DTT to a final volume of 20 μl and incubated at 25°C for 60 min. The reactions were supplemented with 2 μl 25% (w/v) ficoll. Samples were resolved by electrophoresis on a 5% (v/v) polyacrylamide gel containing 0.5× TBE buffer at 150 V for 0.5 h, then 300 V for 2.5 h in 0.5× TBE buffer. Fluorescently labelled DNA oligomers were detected by scanning using a Fujifilm FLA-5100 image reader.

Single turnover kinetic assays

The single turnover assays were based on a previously described protocol (15). Hex-labelled fluorescent DNA (20 nM; sequences 6 and 11 annealed from Supplementary Table S2) was incubated with wild-type and Y89D PrimPol (200 nM). The DNA/PrimPol mixture was incubated with

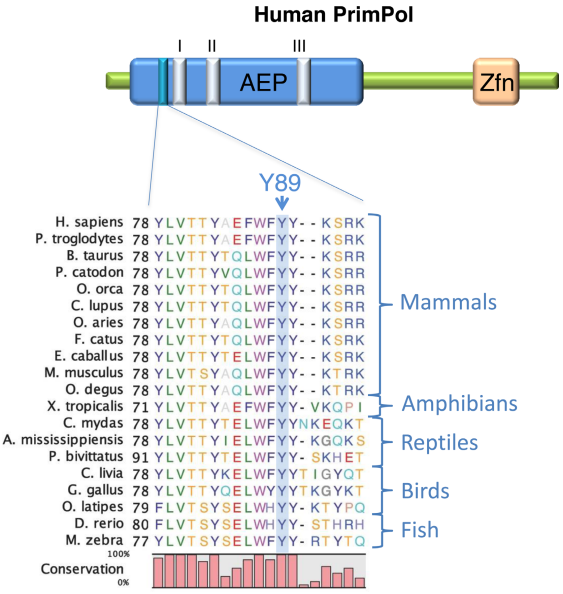


Figure 1. PrimPol^{Y89D} mutation is located adjacent to active site Motif I. An alignment of PrimPol protein sequences from a variety of chordates, including examples of mammals, reptiles, amphibians, birds and fish. The position of the mutated tyrosine 89 in human PrimPol is denoted by an arrow and is labelled 'Y89'. The relative proximity of this residue to the catalytic motifs (I–III) is shown in the AEP domain. The zinc finger domain is denoted by 'Zfn'.

varying concentrations of dATP for time points ranging from 0.5 to 60 min and each reaction was carried out in triplicate. Reactions were quenched by the addition of 2× stop buffer. The products were resolved by electrophoresis as in the primase assays. The concentrations of extended and unextended fluorescent DNA primers were measured using ImageQuant software (GE Life Sciences). The concentration of extended product relative to total fluorescent DNA loaded was plotted as a function of time and the data were fit to a single-exponential curve to obtain the k_{obs} for the different dATP concentrations using the following exponential equation:

$$[product] = A(1 - e^{-k_{obs}t})$$

A secondary plot of k_{obs} as a function of dATP concentration was fitted to the following hyperbolic equation:

$$k_{obs} = \frac{k_{pol}[dATP]}{K_{d(dNTP)} + [dATP]}$$

where k_{pol} is the rate of polymerization and $K_{d(dNTP)}$ is the equilibrium dissociation constant for dATP.

DNA extension processivity assays

The PrimPol variants were analysed for processivity as described previously (16). PrimPol (100 nM) was pre-incubated at 37°C with 60 nM substrate DNA (sequences 5 and 6 annealed in Supplementary Table S2), 10 mM Bis-Tris-Propane-HCl (pH 7.0), 10 mM $MgCl_2$ and 1 mM DTT for 30 min. Reactions were initiated by adding deoxyribonucleotides and excess sonicated herring sperm DNA (1 mg/ml) as an enzyme trap. Reactions were quenched after time points of 15, 30, 60, 120 and 360 s with addition of 2× stop buffer and boiled at 95°C for 5 min. Samples were resolved by electrophoresis as described above.

Circular dichroism (CD) spectroscopy

Samples were buffer exchanged into a buffer of 400 nM NaF, 15 mM Tris-HCl (pH 7.5) and concentrated to 3 mg/ml. The concentration of recombinant protein was determined before and after the CD scans, to ensure accurate secondary structure predictions. Samples were placed in a 0.1 mm quartz cuvette (Starna, Essex, UK) and measurements were taken using a JASCO J-715 spectropolarimeter (JASCO). The CD spectrum of the empty buffer was sub-

tracted from that of the sample and time constant was set to 4 s with a scan rate of 50 nm/min. The bandwidth was 1 nm and the sensitivity set to standard. Scans were performed in triplicate from 320 to 180 nm with a 0.1 nm data pitch and continuous scan mode. A Peltier device was used to maintain a temperature at 20°C.

Protein denaturation curves

The $PrimPol_{1–354}$ and $PrimPol^{Y89D}{}_{1–354}$ constructs were both buffer exchanged into a buffer of 200 mM NaF, 15 mM Tris-HCl (pH7.5). Note that 45 μl of 1 μM protein solution was added to 15 μl of SYPRO Orange, resulting in a final protein concentration of 0.75 μM. A control of 45 μl was also added to 15 μl of SYPRO Orange. 20 μl of each sample was aliquoted in triplicate into a 96-well plate. Protein melting experiments were carried out using the LightCycler 480 System II (Roche). The instrument was configured with a detection format of 465 nm as the wavelength of excitation and 580 nm as the emission wavelength to detect SYPRO Orange-specific signal. Denaturation curve fluorescent signal was acquired within a range of 20°C–80°C using a ramping rate of 0.03°C s^{-1} and an acquisition of 20 data points per degree celsius. Melting temperatures (T_m) were determined through the measurement of the lowest point of the negative differential of the denaturation curve. Data was corrected for the background signal of the buffer conditions.

Complementation and survival assays in PrimPol$^{-/-}$ avian cells

DT40 cells were grown at 39°C in RPMI 1640 medium supplemented with 10 μM β-mercaptoethanol, 1% L-glutamine, 1% penicillin streptomycin, 10% fetal calf serum, 1% chicken serum (Sigma). PrimPol$^{-/-}$ cells were complemented with human $PrimPol^{Y89D}$ cloned into pCI-neo as described previously (14) and expression was confirmed by western blotting alongside wild-type complemented stable cells generated previously (14). DNA replication fork speed and stalling after UV damage was measured by fibre analysis as described previously and carried out in triplicate (4). Cell survival after UV-C damage was measured using CellTiter-blue (Promega). Briefly, 40×10^4 cells (or a serial dilution for decreased damage) were resuspended in a small volume of phosphate buffered saline and exposed to increasing doses of UV-C, media was replaced and cells were grown for 48 h. Cells were then transferred to 96-well plates and viability was measured with CellTiter-blue.

RESULTS

PrimPol^{Y89D} retains DNA primase but loses its RNA primase activity

To determine if PrimPol displays an underlying biochemical defect that contributes to the high myopia clinical phenotype, we first examined the primase activity of the PrimPol Y89D ($PrimPol^{Y89D}$) variant. *De novo* primer synthesis remained largely unchanged between the mutant and the wild-type in the production of DNA primers from dNTPs (Figure 2). However, unlike the wild-type enzyme,

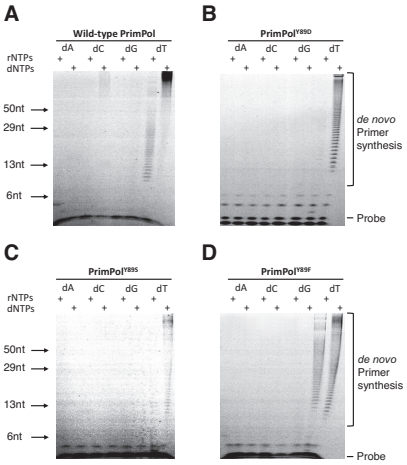


Figure 2. Primase activities of human PrimPol and $PrimPol^{Y89D}$. (A) Human PrimPol has primase activity and can produce *de novo* primers using rNTPs and dNTPs in a homo-polymer template-dependent fashion. (B) $PrimPol^{Y89D}$ partially loses its primase activity. This variant of PrimPol cannot synthesize RNA primers but is an active primase utilizing dNTPs. (C) $PrimPol^{Y89S}$ shows significant loss of the enzymatic activity with residual primase activity observed in presence of dNTPs. (D) $PrimPol^{Y89F}$ is an active primase variant capable of synthesizing both RNA and DNA primers with efficiency comparable to wild-type PrimPol.

$PrimPol^{Y89D}$ was unable to synthesize RNA primers in the presence of even high concentrations of rNTPs. While it has been reported that PrimPol is more proficient at producing DNA over RNA primers *in vivo*, we remained intrigued as to why this mutation in PrimPol conferred a defect in RNA primer synthesis. To determine whether this disruption in activity was a result of the carboxylic acid moiety of the aspartate or was due to the removal of the hydrophobic ring of tyrosine, we produced two additional PrimPol variants, Y89S and Y89F. The serine side-chain of Y89S could discriminate whether the charge of the aspartate caused the altered activity we observed with Y89D while retaining the hydroxyl group of tyrosine, and the phenylalanine of Y89F could help determine if it was the hydrophobic ring of tyrosine that was important for retaining rNTP primase activity. Upon testing Y89S and Y89F for primase activity (Figure 2), we found that the $PrimPol^{Y89S}$ has an activity comparable to Y89D but the Y89F variant of PrimPol returned primase activity back to wild-type, restoring the ability of the enzyme to produce RNA primers.

PrimPol^{Y89D} mutant has limited processivity

PrimPol comprises two intrinsic catalytic activities. To determine any catalytic deficiencies associated with the Y89D

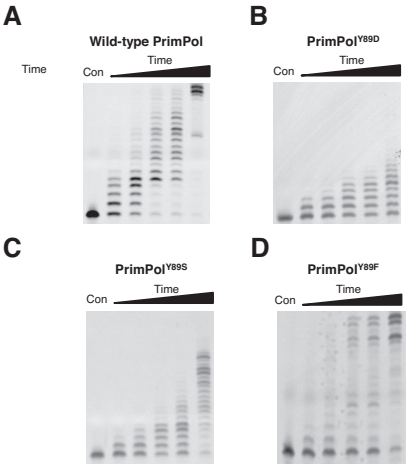


Figure 3. Primer extension activities of human PrimPol and PrimPol^{Y89D}. (A) Human PrimPol has primer extension activity and produces full-length templates following incubation for up to 60 min (lane 1 is a control and lanes 2–6 represent time points at 0.5, 1, 3, 5 and 60 min). (B) PrimPol^{Y89D} shows a marked reduction in polymerase activity, incorporating up to eight nucleotides following incubation for 60 min. (C) PrimPol^{Y89S} also shows a loss in polymerase activity relative to wild-type PrimPol, though it is a more active polymerase than PrimPol^{Y89D}. (D) PrimPol^{Y89F} restores polymerase activity. The Y89F variant can fully extend primers but there is an increased level of unextended primer relative to wild-type PrimPol.

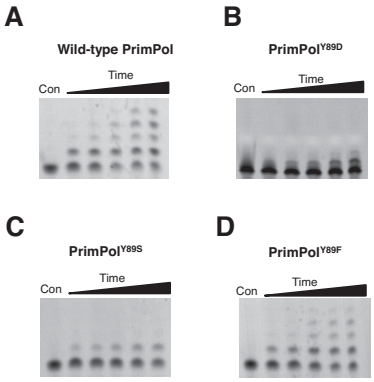


Figure 4. Processivity of human PrimPol and PrimPol^{Y89D}. Primer extension assays were performed in the presence of an excess of herring sperm DNA trap to ensure the polymerase only binds once to the template DNA. Control reactions were also carried out to test the effectiveness of the trap DNA (data not shown). The first lanes represent a no-enzyme control and lanes 2–6 represent time points of 15, 30, 60, 120 and 360 s. (A) Human PrimPol is a poorly processive enzyme, inserting up to four nucleotides opposite a templating DNA strand in a single binding event. (B) PrimPol^{Y89D} has marked reduction in processivity, and is only able to insert a single nucleotide opposite a templating DNA strand per single binding event. (C) PrimPol^{Y89S} can also insert only one single nucleotide opposite a templating DNA strand. (D) PrimPol^{Y89F} restores processivity and has the ability to insert up to four nucleotides, similarly to wild-type PrimPol.

point mutation, we must also consider its polymerase activity in addition to its primase activity (Figure 3). Primer extension assays on PrimPol showed that the polymerase activity of PrimPol^{Y89D} was significantly reduced in comparison to the wild-type enzyme. We next tested if this activity was restored by either Y89S or Y89F variants of PrimPol (Figure 3) and observed that Y89S had slightly higher activity than the Y89D variant but was not restored to wild-type levels. In contrast, Y89F was fully proficient at extending primers opposite a templating strand. Notably, although PrimPol^{Y89F} was able to fully extend primers, and showed full extension at time points as short as 3 min (Figure 3D, lane 4), more primers were left unextended compared with wild-type PrimPol, suggesting that this enzyme has higher activity once it has bound DNA but coordinates it less efficiently.

PrimPol was previously shown to be a distributive DNA polymerase, inserting only up to four nucleotides during a single binding event (Figure 4). We next determined whether the processivity of the mutant was altered, causing the observed reduction in the polymerase activity. PrimPol^{Y89D} variant could only insert a single nucleotide opposite a templating strand of DNA, making it an even more distributive enzyme than wild-type PrimPol (Figure 4). To determine whether the aromatic ring of tyrosine was the main determinant for enzyme activity, we tested the processivity activities of Y89S and Y89F variants. Consistent with Y89D, PrimPol^{Y89S} could only insert a single nucleotide. However, in common with the wild-type enzyme, PrimPol^{Y89F} could insert up to four nucleotides. This establishes that the presence of an aromatic side-chain at this position is requisite for maintaining more processive extension by PrimPol.

Reduced DNA binding contributes to the low processivity of PrimPol^{Y89D}

A reduction in processivity can be attributed to alterations in either dNTP or substrate DNA binding. To test the binding efficiencies of wild-type and PrimPol^{Y89D}, we introduced the mutation into the polymerase domain (PrimPol$_{1–354}$) of PrimPol. This ensured that the contributory binding effects of the zinc finger were removed to enable assessment of the alteration in binding caused solely by this point mutation in the polymerase domain. As the polymerase domain is able to bind both single- and double-stranded DNA (14), we chose to use a labelled overhanging DNA substrate in our EMSAs, similar to the substrates PrimPol would encounter *in vivo*. The binding efficiency of PrimPol$_{1–354}{}^{Y89D}$ to DNA was almost four times lower than

PrimPol$_{1–354}$. The wild-type $K_{D(DNA)}$ was 0.67 μM, compared with Y89D $K_{D(DNA)}$ of 2.55 μM (Figure 5).

PrimPol^{Y89D} has reduced dNTP binding but its fidelity is not altered

To evaluate the effects of this mutation on the enzymatic properties of PrimPol, a kinetic analysis was performed to determine the catalytic efficiency (k_{pol}) and binding constants for dATP ($K_{D(dNTP)}$) opposite a templating thymine (Supplementary Figure S1). We found the wild-type k_{pol} to be 6.98 min^{-1} and the $K_{D(dNTP)}$ to be 15.51 μM (Figure 6). PrimPol is an extremely slow enzyme—for comparison the k_{pol} of base excision repair enzyme DNA polymerase β is 1944 min^{-1} (17). We determined the Y89D variant of PrimPol to be a slightly slower polymerase, with a k_{pol} of 4.07 min^{-1} and binds dNTPs over 10-fold less competently, with a $K_{D(dNTP)}$ of 170.21 μM (Figure 6). The reduction in catalytic efficiency and dramatic reduction in dNTP binding may or may not be dependent on the initial binding of DNA.

If dNTP coordination had been severely affected in the Y89D variant, we would expect quite a deviation from wild-type PrimPol in the fidelity of nucleotide incorporation. However, we found that the Y89D variant of PrimPol did not have any appreciable variation from the wild-type enzyme in terms of fidelity (Supplementary Figure S2). If there was an alteration in the coordination of incoming dNTPs, there would also be a lower probability that this enzyme would be able to bypass the highly distortive lesions that the wild-type enzyme could bypass. However, we observed that PrimPol^{Y89D} was able to synthesize through even highly distortive lesions, such as a 6–4 photoproduct (Supplementary Figure S3).

Aromatic residues often act as a steric gate for sugar discrimination (18). Mutation of these key residues can result in the incorporation of rNTPs instead of dNTPs opposite a templating DNA strand (19). PrimPol^{Y89D}, like wild-type PrimPol, was also unable to extend primers using rNTPs (Supplementary Figure S4). Therefore, it is highly unlikely that this tyrosine residue is involved in discerning the sugar of the incoming nucleotide, particularly as the Y89D variant also loses its ability to prime using rNTPs.

Y89D mutation induces significant structural changes to PrimPol

The observed reduced enzyme activity of PrimPol^{Y89D} could be the result of the loss of aromatic interactions involved in local base stacking events with either the template or the incoming nucleotide. However, it may instead be required for maintaining the structural integrity of the enzyme and mutation of this tyrosine causes more global effects in the structure of PrimPol. To test this second possibility, we subjected the polymerase domains of both the wild-type and PrimPol^{Y89D} to thermal denaturation and CD to check for alterations. There was a striking difference in both the thermal denaturation profiles and the CD spectra of the two enzymes (Figure 7), suggesting this point mutant results in global changes to the enzyme that causes the reduction in activity observed above. The melting point of the polymerase dropped from 49.12°C for the wild-type polymerase domain compared to 42.50°C for PrimPol^{Y89D} (Figure 7A). This significant drop in melting temperature suggests that the enzyme has become less stable and more structurally disordered following the point mutation. Analyses of the secondary structure from the CD spectra of the two variants' polymerase domains by the variable selection algorithm (CDSSTR) (20), which provides better fits for globular proteins, revealed a difference in the α-helical content of these domains (37% for wild-type PrimPol compared with 21% for Y89D), as well as β-strand content (25% for wild-type compared with 29% for Y89D). Additionally, this analysis predicted 30% of the structure of the PrimPol^{Y89D} polymerase domain to be unstructured, whereas just 24% of wild-type was predicted to be unstructured.

PrimPol^{Y89D} significantly decreases DNA replication fork rates *in vivo*

To gain further insights into how the Y89D mutant form of PrimPol may contribute to the myopic phenotype observed in human patients, we studied the influence of PrimPol^{Y89D} on DNA replication in damaged and undamaged cells. We utilized a DT40 PrimPol deleted cell line (PrimPol $^{-/-}$) (4) to examine the ability of PrimPol^{Y89D} to complement this loss, as previously reported for other mutant forms of PrimPol (14). We stably expressed human PrimPol^{Y89D} at similar levels to WT PrimPol in PrimPol knockout cells (Figure 8A), with no obvious effects on cell growth.

We reported previously that PrimPol$^{-/-}$ cells have a greater sensitivity to UV irradiation thus we first examined the effects of increasing doses of UV-C on PrimPol^{Y89D} complemented cell lines. Although PrimPol$^{-/-}$ cells complemented with wild-type PrimPol showed a reversal of this sensitivity, cells complemented with PrimPol^{Y89D} maintained a decreased viability after UV damage (Figure 8B). Next, we measured the replication fork speeds in PrimPol complemented cell lines using DNA fibre analysis, as previously described (4,14). Strikingly, we observed a dramatic decrease in fork speeds in cells complemented with PrimPol^{Y89D} (0.84 kb/min) compared with cells carrying the wild-type PrimPol (1.66 kb/min) (Figure 8C). This decrease in fork speed was significantly greater than that previously reported in PrimPol$^{-/-}$ cells (1.45 kb/min; (4,14)), suggesting the Y89D mutant form of PrimPol may actually impede on-going DNA replication. In contrast, PrimPol^{Y89D} complemented the increased stalling defect observed in PrimPol$^{-/-}$ cells treated with UV-C, equally as well as WT PrimPol (Figure 8D).

DISCUSSION

A recent exome sequencing study has implicated a point mutation (Y89D) in the DNA damage tolerance polymerase, PrimPol, in high myopia in human patients. However, this report did not investigate the molecular consequences of this genetic alteration. Here, we report that the Y89D variant of PrimPol, previously found to be associated with high myopia, is a low processivity variant of PrimPol. Unlike wild-type PrimPol, it inserts just one nucleotide in

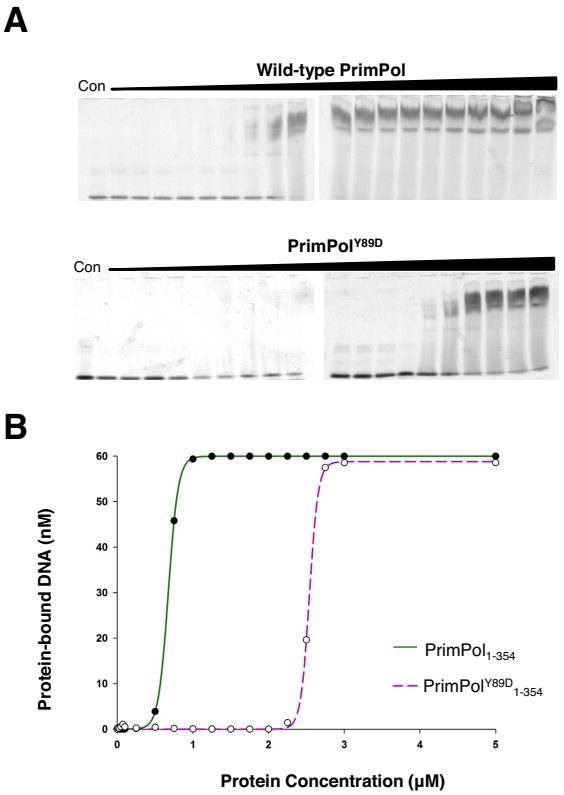


Figure 5. DNA binding efficiency of polymerase domains of wild-type and PrimPol^{Y89D}. (A) The polymerase domains (PrimPol$_{1–354}$) for both wild-type and PrimPol^{Y89D} were used to eliminate the effects of binding from the zinc finger domains. Wild-type PrimPol domain bound to DNA at a much lower concentration than PrimPol^{Y89D} as determined by EMSAs. The concentrations used in both of these EMSAs were 0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 5 and 10 μM. (B) These assays were subsequently quantified to determine $K_{D(DNA)}$ values for the polymerase domains of wild-type PrimPol (solid green line) and its Y89D variant (broken purple line). $K_{D(DNA)}$ of wild-type PrimPol was 0.69 μM, while the $K_{D(DNA)}$ of PrimPol^{Y89D} was 2.55 μM.

a single binding event and has a significantly lower affinity for both dNTPs and DNA. Notably, this mutant did not exhibit any alteration in the fidelity or TLS spectrum relative to wild-type PrimPol. To determine whether this altered activity was a result of losing the hydrophilic hydroxyl moiety or the hydrophobic ring moiety, we constructed two more variants in which tyrosine 89 was mutated to either phenylalanine or serine. The Y89F mutant restored processivity to near-wild-type levels but Y89S had a similar activity to Y89D. In addition, we found that the Y89D PrimPol variant was unable to produce ribonucleotide primers, but this activity was restored in the Y89F variant, indicating that PrimPol requires a hydrophobic aromatic ring at position 89 to maintain processivity but the presence of a hydroxyl moiety is dispensable.

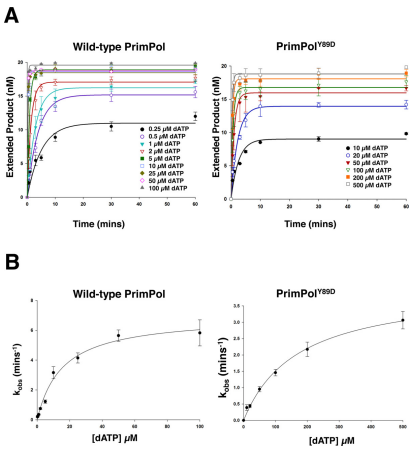


Figure 6. Kinetic analysis of human PrimPol and PrimPol^{Y89D}. Single turnover kinetics of wild-type and PrimPol^{Y89D} to determine catalytic efficiency (k_{pol}) and dNTP binding efficiency ($K_{D(dNTP)}$). (A) The concentration of extended DNA product, as determined by electrophoresis, was plotted against time and fit to an exponential curve for a number of dATP concentrations for wild-type and PrimPol^{Y89D}. These data were fit to exponential curves as described in Equation (1) and k_{obs} was determined. (B) k_{obs} was subsequently plot against dATP concentration for wild-type PrimPol and PrimPol^{Y89D} and these data were fit to hyperbolic curves as described in Equation (2) to determine k_{pol} and $K_{D(dNTP)}$. k_{pol} for wild-type PrimPol was found to be 6.98 $min^{-1} \pm 0.40$ and $K_{D(dNTP)}$ was 15.51 $\mu M \pm 2.71$. k_{pol} for PrimPol^{Y89D} was 4.07 $min^{-1} \pm 0.19$ and $K_{D(dNTP)}$ was 170.21 $\mu M \pm 18.80$.

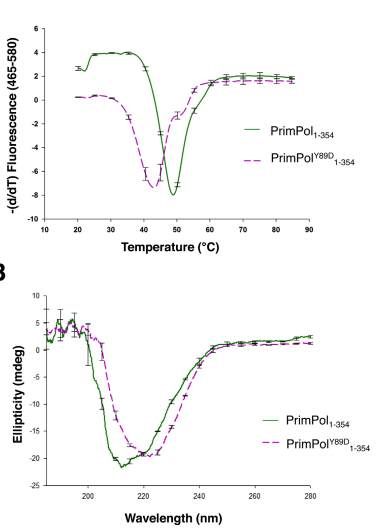


Figure 7. Y89A induces a structural perturbation in PrimPol. The polymerase domain of the Y89D variant of PrimPol has a different global structure and thermal stability compared to wild-type PrimPol. (A) Thermal denaturation of the polymerase domain of PrimPol and its Y89D variant. The negative differential of the fluorescence profiles of the polymerase domains (PrimPol$_{1-354}$) of wild-type PrimPol (solid green line) and its Y89D (broken purple line) variant plotted against temperature to determine their melting points. The melting point of the polymerase domain of the wild-type enzyme was found to be 49.12°C, while the Y89D variant had a considerably lower melting point at 42.50°C. (B) CD spectra of the polymerase domains of wild-type PrimPol and PrimPol^{Y89D} show a substantial difference in the global structure of PrimPol^{Y89D} relative to the wild-type enzyme.

Insights from structural studies on primases and polymerases suggest two different potential explanations for the reduction in polymerase processivity observed in this study. The conserved tyrosine at position 89 could have an important role in base stacking or could be required for the structural integrity of the protein. There is an invariant phenylalanine in nonhomologous end-joining (NHEJ) AEPs that plays a vital role in the recognition and orientation of both the templating and the incoming bases through stacking interactions with both bases and is responsible for moving the incoming nucleotide into the enzyme's active site (21,22). Phenylalanine 290 in human Pol η helps form the hydrophobic core of the thumb domain (23). Mutation of this residue is associated with the genetic disorder *xeroderma pigmentosum variant*, which is characterized by a susceptibility to skin cancers caused by a failure to perform proficient TLS opposite UV lesions. This mutation negates polymerase activity and is predicted to destabilize the thumb domain, resulting in reduced primer binding. While the structure of PrimPol has not yet been elucidated, it is conceivable that the tyrosine at position 89 is involved in preserving the structure of a domain analogous to the thumb domain that is found in other polymerases involved in binding/coordinating the templating DNA strand. Tyrosine 89 is situated in a conserved hydrophobic pocket containing other conserved aromatic residues including Y87, F88 and Y90, as well as a hydrophobic residue at position 86 (Figure 1) and it is likely that disruption of Y89 disrupts a conserved interaction between these amino acids. A secondary structure prediction from PSIPRED (24) predicts that all of these amino acids lie within an α-helical region and the Y89 mutation may disrupt this helix. The CD data certainly supports this model, indicating that the Y89D variant has less α-helical structure than the wild-type. We postulate that the Y89D mutation causes a disruption in the global structure of the polymerase domain of PrimPol. As the affinity for DNA and dNTPs is decreased in the Y89D variant, but the apparent fidelity of the enzyme remains unchanged, we suggest that dNTP binding is depen-

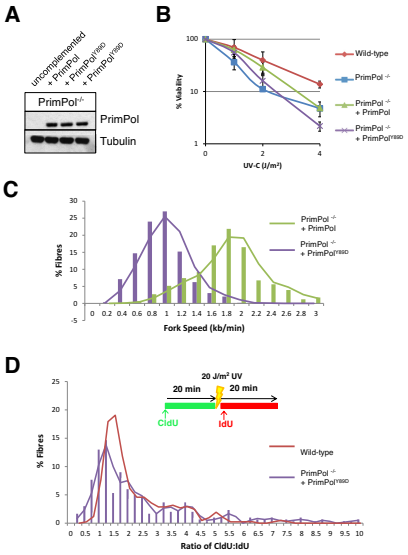


Figure 8. PrimPol Y89D cannot fully complement the UV sensitivity seen in PrimPol$^{-/-}$ DT40 cells due to decreased replication fork speeds. (A) Western blot showing overexpression of wild type and Y89D point mutant forms of PrimPol in PrimPol knockout (PrimPol$^{-/-}$) DT40 chicken cells. (B) Cell viability was measured using CellTiter-Blue 48 h after increasing doses of UV-C were administered to the different cell types. Results represent three independent repeats and the error bars represent the standard deviation across the repeats. (C) DNA fibres were measured after sequential labelling with CldU and IdU for 20 min each to calculate replication speed. Results represent three independent repeats. (D) A 20 J/m^2 UV-C pulse was included between the two labels and the ratio of the pre- and post-UV-C labels was calculated, with increased ratios representing replication stalling after UV. Results represent three independent repeats.

dent on DNA binding, which is disrupted by the structural changes induced by this mutation.

In vivo analysis identified a dramatic decrease in replication fork rates in cells expressing PrimPol^{Y89D} and, despite being able to complement fork stalling in PrimPol$^{-/-}$ cells after UV damage, it was unable to prevent decreased cell viability. Consistent with *in vitro* assays, PrimPol^{Y89D} causes decreased replication fork rates in cells and it appears that this defect is enough to decrease the damage tolerance of these cells. Both our *in vivo* and *in vitro* studies on this mutant indicate that although PrimPol^{Y89D} retains the ability to bypass lesions, its capacity to replicate DNA is severely compromised. Therefore, it appears that although PrimPol^{Y89D} is recruited to stalled replisomes, the underlying replication defect causes a significant slowing of replication forks, potentially resulting in incomplete lesion bypass, causing cells to progress through the cell cycle with underreplicated genomic DNA. Clearly, these cells carry an abundance of PrimPol compared to wild-type levels and thus replication fork speed alone in endogenous mutants may not be so significantly altered. However, in cases where PrimPol is actively recruited to on-going forks or damage sites, this 'slow' mutant is likely to cause significant problems. Therefore, unexpectedly, it appears that harbouring the PrimPol^{Y89D} mutation is more detrimental than not having PrimPol at all. This data correlates with the phenotypes observed in patients carrying this mutation as problems are observed in tissues where levels of UV damage are expected to be high. Thus, in tissues where PrimPol is recruited in higher abundance to replicate through DNA damage, such as eyes, cellular defects are much more likely to be prevalent.

The link between reduced PrimPol polymerase activity and myopigenicity remains unclear. Previous studies into the genes that predispose high myopia have identified two clusters of mutations, one in the outer retina affecting photoreceptors and ON-bipolar cell function and another in the scleral extracellular matrix composition and metabolism (11). Significantly, mutations in PrimPol do not fit into either category. The severity of the clinical phenotype of patients with the Y89D mutation varied (9) and it is highly likely that this gene is not a singular, unique aetiological factor and that other genetic and environmental factors also contribute to pathogenesis. Although replication defects associated with PrimPol^{Y89D} suggests that it contributes to the reported clinical phenotypes, further work is required to establish if a direct link exists between replication stress and onset of high myopia.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors would like to thank Prof. Shigenori Iwai who kindly provided DNA containing a 6–4 photoproduct.

FUNDING

BBSRC [BB/J018643/1, BB/H019723/1 and BB/M004236/1 to A.J.D.]; MRC [G0080130]. Funding for open access charge: RCUK.
Conflict of interest statement. None declared.

REFERENCES

1. Aguilera,A. and Gómez-González,B. (2008) Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.*, **9**, 204–217.
2. Li,X. and Heyer,W.-D. (2008) Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.*, **18**, 99–113.
3. Sale,J.E., Lehmann,A.R. and Woodgate,R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.*, **13**, 141–152.
4. Bianchi,J., Rudd,S.G., Jozwiakowski,S.K., Bailey,L.J., Soura,V., Taylor,E., Stevanovic,I., Green,A.J., Stacker,T.H., Lindsay,H.D. *et al.* (2013) Eukaryotic PrimPol bypasses UV photoproducts during chromosomal DNA replication. *Mol. Cell*, **52**, 566–573.

5. Rudd,S.G., Glover,L., Jozwiakowski,S.K., Horn,D. and Doherty,A.J. (2013) PPL2 translesion polymerase is essential for the completion of chromosomal DNA replication in the African Trypanosome. *Mol. Cell*, **52**, 554–565.
6. Wan,L., Lou,J., Xia,Y., Su,B., Liu,T., Cui,J., Sun,Y., Lou,H. and Huang,J. (2013) hPrimpol1/CCDC111 is a human DNA primase-polymerase required for the maintenance of genome integrity. *EMBO Rep.*, **14**, 1104–1112.
7. Mourón,S., Rodriguez-Acebes,S., Martínez-Jiménez,M.I., Garcia-Gomez,S., Chocron,S., Blanco,L. and Mendez,J. (2013) Repriming of DNA synthesis at stalled replication forks by human PrimPol. *Nat. Struct. Mol. Biol.*, **20**, 1383–1389.
8. Garcia-Gomez,S., Reyes,A., Martínez-Jiménez,M.I., Chocrón,E.S., Mourón,S., Terrados,G., Powell,C., Salido,E., Mendez,J., Holt,I.J. *et al.* (2013) PrimPol, an archaic primase/polymerase operating in human cells. *Mol. Cell*, **52**, 541–553.
9. Zhao,F., Wu,J., Xue,A., Su,Y., Wang,X., Lu,X., Zhou,Z., Qu,J. and Zhou,X. (2013) Exome sequencing reveals CCDC111 mutation associated with high myopia. *Hum. Genet.*, **132**, 913–921.
10. Vitale,S., Sperduto,R.D. and Ferris,F.L. (2009) Increased prevalence of myopia in the United States between 1971–1972 and 1999–2004. *Arch. Ophthalmol.*, **127**, 1632–1639.
11. Morgan,I.G., Ohno-Matsui,K. and Saw,S.M. (2012) Myopia. *Lancet*, **379**, 1739–1748.
12. Saw,S.M., Gazzard,G., Shih-Yen,E.C. and Chua,W.H. (2005) Myopia and associated pathological complications. *Oph. Phys. Optics*, **25**, 381–391.
13. Iyer,L.M., Koonin,E.V., Leipe,D.D. and Aravind,L. (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.*, **33**, 3875–3896.
14. Keen,B.A., Jozwiakowski,S.K., Bailey,L.J., Bianchi,J. and Doherty,A.J. (2014) Molecular dissection of the domain architecture and catalytic activities of human PrimPol. *Nucleic Acids Res.*, **42**, 5830–5845.
15. Nemec,A.A., Donigan,K.A., Murphy,D.L., Jaeger,J. and Sweasy,J.B. (2012) Colon cancer-associated DNA polymerase β variant induces genomic instability and cellular transformation. *J. Biol. Chem.*, **287**, 23840–23849.
16. Washington,M.T., Johnson,R.E., Prakash,S. and Prakash,L. (1999) Fidelity and processivity of Saccharomyces cerevisiae DNA polymerase eta. *J. Biol. Chem.*, **274**, 36835–36838.
17. Nemec,A.A., Murphy,D.L., Donigan,K.A. and Sweasy,J.B. (2014) The S229L colon tumor-associated variant of DNA polymerase beta induces cellular transformation as a result of decreased polymerization efficiency. *J. Biol. Chem.*, **289**, 13708–13716.
18. Brown,J.A. and Suo,Z. (2011) Unlocking the sugar 'Steric Gate' of DNA polymerases. *Biochemistry*, **50**, 1135–1142.
19. Donigan,K.A., McLenigan,M.P., Yang,W., Goodman,M.F. and Woodgate,R. (2014) The steric gate of DNA polymerase ι regulates ribonucleotide incorporation and deoxyribonucleotide fidelity. *J. Biol. Chem.*, **289**, 9136–9145.
20. Lobley,A., Whitmore,L. and Wallace,B.A. (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*, **18**, 211–212.
21. Pitcher,R.S., Brissett,N.C., Picher,A.J., Andrade,P., Juarez,R., Thompson,D., Fox,G.C., Blanco,L. and Doherty,A.J. (2007) Structure and function of a mycobacterial NHEJ DNA repair polymerase. *J. Mol. Biol.*, **366**, 391–405.
22. Brissett,N.C., Martin,M.J., Pitcher,R.S., Bianchi,J., Juarez,R., Green,A.J., Fox,G.C., Blanco,L. and Doherty,A.J. (2011) Structure of a preternary complex involving a prokaryotic NHEJ DNA polymerase. *Mol. Cell*, **41**, 221–231.
23. Biertümpfel,C., Zhao,Y., Kondo,Y., Ramón-Maiques,S., Gregory,M., Lee,J.Y., Masutani,C., Lehmann,A.R., Hanaoka,F. and Yang,W. (2010) Structure and mechanism of human DNA polymerase η. *Nature*, **465**, 1044–1048.
24. Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.